NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

VISION-BASED INTEREST POINT EXTRACTION EVALUATION IN MULTIPLE ENVIRONMENTS

by

Zachary Dean McKeehan

September 2008

Thesis Co-Advisors: Mathias Kölsch
Kevin Squire

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 2008	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE Vision-Based Interest Point Extraction Evaluation in Multiple Environments	5. FUNDING NUMBERS
6. AUTHOR(S) Lieutenant Zachary Dean McKeehan	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited	12b. DISTRIBUTION CODE A

13. ABSTRACT (maximum 200 words)

Computer-based vision is becoming a primary sensor mechanism in many facets of real world 2-D and 3-D applications, including autonomous robotics, augmented reality, object recognition, motion tracking, and biometrics. Vision's ability to utilize non-volatile features to serve as permanent landmarks in motion tracking provides a superior basis for applications such as initial self-localization, future re-localization, and 3-D scene reconstruction and mapping. Furthermore, the increased reliance of the United States armed forces on the standoff war-fighting capabilities of unmanned and autonomous vehicles (UXV) in, on, and above the sea, necessitates better overall navigation capabilities of these platforms. Towards this end, we draw upon existing technology to measure and compare current visual interest point extractor performance. We utilize an inventory of interest point extractors to define and track interest points through physical transformations captured in images of various scene classifications. We then perform a preliminary determination of the best-suited extraction descriptor for each visual scene given multi-frame interest point persistence with maximum viewpoint invariance. Our research contributes an important cornerstone towards the validation of precision, vision-based navigation, thereby increasing UXV performance and strengthening the security of the United States and her allies worldwide.

14. SUBJECT TERMS Simultaneous Localization and Mapping, SLAM, Epipolar Geometry, Fundamental Matrix, Camera Motion, Vision, Feature Extraction, Interest Point, Feature Detections, Feature Description, Scale Invariant Feature Transform, SIFT	15. NUMBER OF PAGES 207
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

VISION-BASED INTEREST POINT EXTRACTION EVALUATION IN MULTIPLE ENVIRONMENTS

Zachary D. McKeehan
Lieutenant, United States Navy
B.S., Chapman University, 1999

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
September 2008

Author: Zachary Dean McKeehan

Approved by: Dr. Mathias Kölsch
Thesis Co-Advisor

Dr. Kevin Squire
Thesis Co-Advisor

Dr. Peter Denning
Chairman, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Computer-based vision is becoming a primary sensor mechanism in many facets of real world 2-D and 3-D applications, including autonomous robotics, augmented reality, object recognition, motion tracking, and biometrics. Vision's ability to utilize non-volatile features to serve as permanent landmarks in motion tracking provides a superior basis for applications such as initial self-localization, future re-localization, and 3-D scene reconstruction and mapping. Furthermore, the increased reliance of the United States armed forces on the standoff war-fighting capabilities of unmanned and autonomous vehicles (UXV) in, on, and above the sea, necessitates better overall navigation capabilities of these platforms. Towards this end, we draw upon existing technology to measure and compare current visual interest point extractor performance. We utilize an inventory of extractors to define and track interest points through physical transformations captured in images of various scene classifications. We then perform a preliminary determination of the best-suited extraction descriptor for each visual scene given multi-frame interest point persistence with maximum viewpoint invariance. Our research contributes an important cornerstone towards the validation of precision, vision-based navigation, thereby increasing UXV performance and strengthening the security of the United States and her allies worldwide.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Consider the ease with which we humans can determine our position relative to our surroundings and then accurately update that position through our subsequent motion. Medical and physiological research performed in [1], [2] has shown that this phenomenon is a direct result of the remarkably accurate human environment sensing and processing ability. This process of temporal and spatial motion perception appears to be a trivial exercise only because we are not fully conscious of the complex procedure that we take for granted and yet perform flawlessly on a continuous basis.

Let us picture ourselves in a room that has four walls and a door. Imagine that we visually scan the room and notice that there is a door about 12 feet directly across the room. We also notice a small table, a desk and a chair in the room between the door and us. This is a common situation and we would be confident in our ability to walk across the room, avoiding collisions with the small table, the desk and the chair, and proceed to the door and out of the room. We clearly could exit the space with minimal cognitive effort.

Let us now imagine the same scenario except with a small modification: the removal of just one of our senses, vision. We are in the same room except that this time, after a brief visual scan of the room, the lights extinguish, leaving the room completely devoid of light. Now if we desired to leave the darkened room, would we still be confident in our ability to navigate around the obstacles in our quest for the door? Consider that in this situation, as with a lighted room scenario, at the very moment we commence our traversal, neurons in our brain would fire causing the routing of electrical impulses through our nervous system to muscles in our legs. The muscles would respond to the stimulation and contract and extend in autonomic harmony to produce bi-pedal locomotion. We would be unconsciously aware of our body mechanics and the physical kinetics produced by the muscles through a complex sensory feedback system that was

developed and fine-tuned during very early childhood development. This odometery feedback system provides us with a muscle-based perception of distance and direction traversed. We would also sense the echo returns of sound waves generated by our movement as the pressure waves reach our ears from the reflection and refraction by the room and objects. These forms of environmental feedback provide us two pieces of information concerning our movement, but interestingly enough they alone are not sufficient for accurate motion perception. Even if we begin with a precise image in our mind of the room and the obstacle layout, without vision to verify our motion, the additive inaccuracies generated with every step through our muscle-based odometry will reduce our confidence in a perceived self-generated location. This is evident by the instinctive reflex we have to place our arms out in front of us in an exploratory attempt to detect objects prior to collision when we are unable to employ our sense of vision.

Finally, let us consider that instead of a human traversing the room, we are interested in performing a successful traversal with an autonomous, robotic system. Prior research in mechanical robotics has shown that we can successfully implement any number of movement contrivances, such as mechanisms that are ground-based with wheels, tracks, bi-pedal limbs, quad-pedal limbs, or even with fixed-wing and rotary wing mechanisms of flight. These assemblies will enable physical locomotion and would allow the system to maneuver in 3-D space through successive changes in the system's pose and position. We can also implement an electro-mechanical sensory feedback system to provide onboard odometry, measuring system displacement along three axes as a result of motion. As with the human perception of motion, as sensed only through muscle odometry, a robotic odometery system will contain a measure of error in the form of noise. This noise is mainly a result of slippage and it builds over time significantly affecting the accuracy and confidence in a self-localization position for a robotic system [3]. As a result, we need to update and reconcile the onboard odometry with other sensors. Passive and active

sensors are the typical devices used to reconcile odometry. An example of an active sensor is a RADAR or a Light Detection and Ranging (LIDAR) system. The highly accurate range measurements these devices provide are used to update a system-generated odometry to more accurately reflect the resultant pose and position following movement. Active sensors work well for many problem domains; however, due to their expensive price, large package size, high power requirements, and in a military application, the detectability of the ranging energy emissions, other problem domains require a passive sensor. A passive sensor does not emit energy into the environment to perform a measurement, but instead uses energy already available, such as magnetism or light. Unfortunately, devices that measure magnetism suffer from many of the same drawbacks as the active sensors; however, devices that operate in the light domain such as cameras are comparably inexpensive, can be very small and require very minimal power. With this in mind, various research efforts [4], [5], [6], [7], [8], [9] have focused on achieving a human-like perception of motion by a computational system through live digital cameras and computer vision.

Computer-based vision is becoming a primary sensor mechanism in many facets of real world 2-D and 3-D applications, including autonomous robotics, augmented reality, object recognition, motion tracking, and biometrics. Vision's ability to utilize non-volatile features to serve as permanent landmarks in motion tracking provides a superior basis for initial self-localization and future re-localization. Through Computer Vision, we can capture unique interest points in a scene and track their spatial location through successive scene frames. The movement of each feature provides for a model of motion perception. A keystone in this research area is the interest point detection, selection, classification, registration, storage and correlation.

In computer vision, research has shown that for scene classifications, some feature extractors work better than others [10], [11], [12], [13]. However, current vision-based applications do not attempt to select interest point extraction algorithms based on a quantifiable measure of potential performance for the

given scene environment classification. The unbiased employment of interest points can lead to an extraordinarily excessive expenditure of processing power and computational time on what will prove to be largely non-usable data within the problem domain. Today's challenging computer-based vision scenarios are pushing the limits of real-time processing, requiring a judicious and efficient use of processing power.

A. MOTIVATION AND BENEFITS

Ultimately, we would like to create a system capable of perceiving motion with six degrees of freedom (6-DOF) for autonomous rotary-wing aerial robots to conduct Simultaneous Localization and Mapping (SLAM) using only computer vision. We assume that in order to perform real-time aerial SLAM, a robot needs to be equipped with the most efficient mechanism to detect, measure and catalog its surroundings. To enable this capability, this thesis investigates and quantifies how suitable the various feature extractors are in a certain environment. Ultimately, we would like to employ different feature extractors in different parts of a single image.

The immediate benefit of this study is the ability to provide interest point extractor selection based on environment suitability for persistence and viewpoint invariance. This approach can be applied virtually to all feature-based camera-tracking algorithms to sensibly extract interest points, allowing processing resources and algorithms to be applied only to the highest quality points in the most computationally efficient manner.

B. SCOPE OF THESIS

This thesis does not seek to create a new extraction algorithm or to create a new application for their employment, but instead we desire to better employ the ones that already exist. Towards this end, we draw upon existing technology and application frameworks to measure and compare current extractor performance. We utilize an inventory of interest point extractors to define and

track interest points through identical physical transformations in the environments of various scene classifications. We then determine the best-suited extractor for each visual scene given multi-frame interest point persistence with maximum viewpoint invariance. The primary metrics for extractor performance are consistent with previous work in this area [10], [12], [13], [14] and are explained in detail in Chapter III

C. CHAPTER ORGANIZATION

The remainder of this thesis is organized as follows. Chapter II provides the reader with an introduction to the theory behind computer vision and interest point extraction. We also explore recent work and the current state of the art to include previous performance evaluations and employment in SLAM experiments. Chapter III describes the methodology that we used to conduct our research. Chapter IV provides the results of the actual experiments. The final chapter of the thesis, Chapters V, gives a general summary of our work and conclusions and explores the opportunities for future work.

THIS PAGE INTENTIONALLY LEFT BLANK

II. VISION-BASED FEATURE TRACKING AND APPLICATIONS

In visual localization and trajectory tracking, the relative motion of a camera is determined through the movement of objects and background within the camera field of view. The salient needs typically include a motion-tracking system using a camera, and a methodology for obtaining camera trajectories (e.g., [5], [6]). This chapter explores recent work in this field, discusses related concepts, and introduces pertinent terminology and the basic mechanics of visual motion perception and its application in computer vision. We will discuss visual motion perception in humans and computers and then describe the particular aspects of the feature extraction and tracking process in a vision system. We will also explore other research efforts focused along these lines.

A. VISUAL MOTION PERCEPTION

As we described in Chapter I, the biological and psychological computation of motion in the human brain involves tactile, auditory and visual sensory information. Vision and visual motion perception is by far the overriding sensory input.

According to Itan [15], to perceive the motion of an object we must first identify it, note its position, and later identify the same object again, noting its new position. The velocity of the movement is then computed by the change of position divided by the time, as $\delta s / \delta t$. This process is the basis for "correspondence" motion perception models, which function by matching "things" through time.

From a different perspective, we can also talk about the perception of our own motion. Described in detail by S. Coren, et al. [1], a human automatically perceives ego-motion primarily through his or her visual system. As we move forward, our world is visually captured in a radially expanding pattern from the center of our visual field and laterally translated periphery. The change over time

of this flow of stimuli produces a structure and event correspondence called streaming perspective that forms the basis for our perception of motion. The center of this outward flow, called the focus of expansion [1], indicates the direction of our movement.

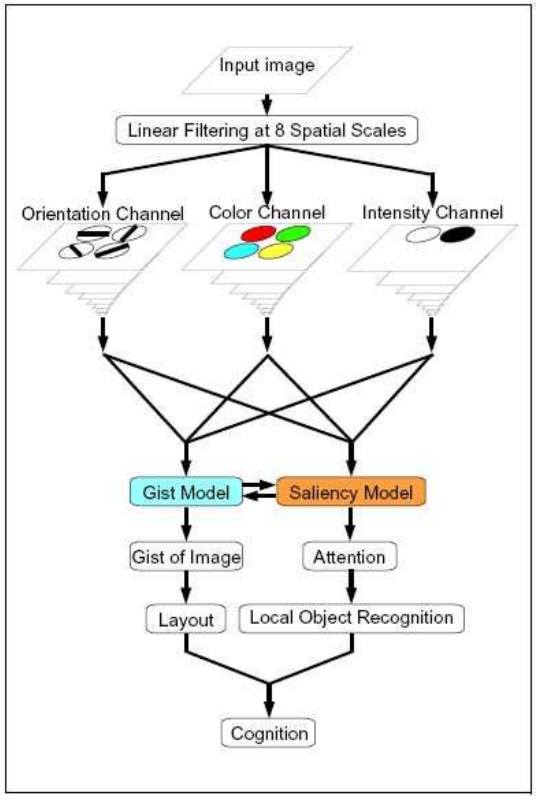


Figure 1. Model of Human Vision with Gist and Saliency [From [16]]

Historically, much research has focused on the problem of replicating a human-like perception of ego-motion using computational devices. More recently, C. Siagian and L. Itti [16] have sought to replicate this form of human perception as a Gist and Saliency model illustrated in Figure 1. The main problem in reproducing the human biological method is understanding exactly how the focus of expansion and flow of stimuli is autonomously processed to produce a perception of ego-motion.

In our research, we seek to enable a more accurate and complete visual sensory input to such a system through the existing hardware configuration. We

believe that our work will lead to a substantial increase in system operation by improving upon exactly how the images are processed in the intermediate steps from capture to motion determination.

B. VISION SYSTEMS

While computers do not yet have eyes as we know them, they can be equipped with cameras that capture digital images. Digital images are sampled from the physical world and are essentially captured projections of visible light. The most common type of camera used in computer vision is a charge coupled device (CCD) camera. A CCD camera uses a small, rectangular piece of silicon to receive and measure incoming light. The CCD wafer is a solid-state electronic component, segmented into an array of individual light-sensitive cells called "photosites." In a typical low-end CCD camera, the light spectrum is sampled and represented in color by overlaying color filters on the photosites. Predominantly a three color filter known as a Bayer filter is used. Four filter areas make up each pixel: one red, one blue and two green, corresponding to the sensitivity curves of the color receptive cones found in the human eye. At the high-end of CCD cameras is a 3-wafer CCD camera. Instead of determining color value by overlaying a filter, a 3-wafer camera employs a beam-splitter prism that separates the continuous light sample into separate color channels and funnels each channel to a separate CCD. This allows for a more precise sampling of the light spectrum for capturing the image.

Continuous sequences of images produced by a CCD camera typically provide the primary input to vision system applications. Each image is processed according to the intended application task, however most systems require an implementation of some form for tracking events or objects between consecutive images in the sequence. This thesis is primarily concerned with the form of vision tracking described in the next section.

C. FEATURE TRACKING

In the process of producing a perception of ego-motion in a computing system, we can sample the visual spectrum of our environment through a CCD camera as described above. The resultant images represent a sequence of scenes with a flow, not unlike the flow of stimuli that produces a streaming perspective in human motion perception. To interpret the sequence of images for optical flow, the computing system must somehow track salient aspects of the scene. Towards this end, real-time vision tracking applications seek to segment a scene into semantically relevant elements in the form of foreground objects and background [14]. The process of segmentation, in theory, allows for pinpoint localization of an object attribute down to the sub-pixel level. This level of accuracy provides a basis for a highly efficient employment of only the most accurate attributes for correspondence. Unfortunately, automatic scene segmentation is a difficult problem and even if the accuracy was sufficient, the current technology is too slow to implement in a real-time system. Carson, et al. [17], implemented Blobworld, the best-known example of robust image segmentation in which the scenes are segmented based on color and texture. Each segment is then matched by shape to a representative object in a database lookup. This implementation performed accurate segmentation with recall scores between .2 and .3, at 5 to 7 minutes per image on a 300MHz Pentium II processor. See [17] for additional Blobworld details. Even with today's processing speeds, the significantly long processing time and required database domain knowledge present challenges for use in an online vision tracking system.

Recent tracking efforts have shifted from foreground-background segmentation methods to the use of local object attributes known as features. The term "feature" is an abstract term that refers to any artifact or region in an image and its unique description. Examples of these features include corner-based artifacts like Harris [18], and Föstner points [19], Difference of Gaussian points (DOG) [20], scale space blobs [21] and Maximally Stable Extremal

Regions (MSER) [22]. As described by the research efforts listed above, there are a great number of ways of finding these artifacts and describing them. Regardless of the specific methods, these artifacts remain projections of object attributes in 3-D space onto a 2-D image plane and therefore shall be referred to in this thesis as interest points.

At the core of many tracking efforts utilizing computer vision, is the fundamental process of matching interest points from one image to another of the same scene. These efforts include vision-based SLAM [4], object or environment modeling [5], geographic registration of aerial imagery, augmented reality for training and combat, autonomous & unmanned system navigation, system control and data processing. Interest point based matching methods are categorized as wide-baseline or short-baseline (also known as narrow-baseline). In short-baseline matching, the interest points within the images of the scene are only expected to change position a short distance from frame to frame, and can therefore be tracked using filtering techniques such as Kalman filters [23]. Their appearance is not expected to change much if at all, particularly their scale and orientation usually remains unchanged (rigid translation only). In wide-baseline matching, the interest points can have great and wide changes in appearance, and must be matched by some other means. While both are important, our work focuses on the more complex process of wide-baseline matching.

Finding a wide-baseline correspondence from one image to another is accomplished through three basic steps. First, we must detect interest points within each image by scanning and detecting distinctive regions such as gradient edges, corners, or unique blobs. This is done because interest point calculation of feature descriptors is computationally expensive and cannot be done for every possible location (every pixel). The interest point detector is crafted in such a manner as to maximize the likelihood that the detected interest points are distinctive and can be reliably found repeatedly in images taken from different vantage points. The next step of wide-baseline tracking aims at uniquely

describing the immediate region surrounding the interest point using a feature vector. This is accomplished by sampling the pixels surrounding the interest point, processing them in some fashion, and producing a vector-based representation that is robustly tolerant to noise and deformations. The combined actions of detection and description produce an invariant feature vector and shall be referred to in this thesis jointly as interest point (or feature) extraction. We also define interest point to represent the center pixel of a detected point or region and feature to represent the interest point as described by a feature descriptor; however, their usage within this thesis is interchangeable. Lastly, the sets of generated interest points are compared between two images to find matches. Matches are found based on a mathematical Euclidean or Mahalanobis distance between the feature vectors. We describe details of these processes below.

1. Interest Points

Interest points capture and describe unique attributes of a region, centered on an attribute of a scene. As an abstract data structure, interest points can be used for computations in applications such as image mosaics [24], video data mining [25], object recognition [26], [27], autonomous vehicle localization [4], [5], [6], [9], texture recognition [28], and image matching [29]. Points of interest are extracted and described in many different ways. Typically, a specific representation is chosen to fit the expected problem domain. In some implementations, interest points can be defined by physical attributes that are exhibited uniquely by an object of interest in the problem space. For example, Harrell et al. [30] used the centroid and diameter of circles to define the features used to track fruit for robotic harvesting. An interest point can also be as simple as a corner in a 2-D image described by the defining intensity gradients as employed by Saeedi, et al. [5]. On the other hand, it can be as complex as Lowe's [31] Scale Invariant Feature Transform (SIFT), a high-level scale invariant

descriptor complete with a 128 dimensional vector computed from the spatial distribution of image gradients over a circular region. We discuss SIFT in detail later in this chapter.

We would like interest points to correspond to semantically meaningful object attributes. However, this is infeasible for most applications, as this would require a high-level interpretation of the scene content. Instead, detectors select local interest points directly based on the underlying intensity patterns of a point or region. In some applications, it is also highly desirable to generate interest points that are permanent and easily detectable. Saeedi, et al. [5] defines these long-term interest points as landmarks. Landmark-based methods perform motion tracking by detecting landmarks in the environment and estimate camera position based on triangulation. These methods must either utilize prepositioned landmarks or learn a local distinctive pattern suitable for tracking during a training phase of deployment. The MINERVA [32] tour guide robot navigates in this fashion and uses the distinct pattern in the ceiling of Smithsonian's National Museum of American History as a mosaic-based landmark template to localize and estimated motion. Systems that localize based on this approach usually require an *a priori* map of the domain space. On the other hand, natural interest point-based designs seek to capture and track naturally occurring object attributes that are extracted directly from the environment [25]. The relative changes in the interest points provide the mechanism the ability to estimate the camera (and hence robot) trajectory and motion.

Another important characteristic of interest points is invariance. When large out-of-plane transformations[1] are expected in the scene, the best approach is to model the appearance changes as mathematical point transformations, and then develop methods for interest point detection that are unaffected by viewpoint changes [14]. Types of desired invariance, or covariance include

[1] An in-plane transformation occurs when the camera motion is such that 3-D objects maintain an undistorted 2-D appearance in the image plane. In an out-of plane transformation, the camera moves out of the 2D plane relative to the object, distorting the captured 2-D appearance.

translation, rotation, illumination, scale and affine (or viewpoint) transformation invariance [10]. The concept of transformation invariance is straightforward for translation, rotation and illumination. Intuitively, we want to be able to find the same interest point even if due to camera movement, it undergoes a rotation transformation, is located in a different area, or is illuminated differently in subsequent frames. Scale invariance involves accounting for changes in overall scale. T. Tuytelaars and K. Mikolajczyk suggest that a detector is considered scale invariant if it provides a reliable match at least up to a scale factor of four [14]. Affine transformations are generalizations on scale transformations. A scale transformation can be non-uniform (anisotropic) and actually affect regions of the image differently in each direction. The non-uniform scaling changes the shape of the image and thus the shape of interest points. A detector that is crafted to only handle scale invariance would not be able to correctly match an interest point that has undergone a significant out-of-plane and affine transformation.

2. Interest Point Detection

Klippenstein and H. Zhang [13] describe interest point extraction as a two-part process of first detecting an interesting attribute and then capturing the attribute as an interest point in a unique, comparable abstract form known as a descriptor. Interest point detectors find the interesting characteristics in a scene image that meet the appropriate property criteria as defined below for the intended application. Recently Mikolajczyk et al., analyzed a large inventory of interest point detectors [10] and descriptors [11] under varying conditions and one of their findings was that the selection of a point detector was less significant than the selection of a descriptor. Their evaluation techniques will discussed in detail later in this chapter.

There are many ways to detect points of interest within an image, most of which can be sectioned into three categories defined by Schmidt, et al. in [33], as contour based, intensity based and parametric model based. In a contour-based

approach, a detector will first extract the contours of the scene and then use the inflection points or polygonal intersections as the interest point locations. Intensity based methods seek out regions of high intensity gradients as interest points. Parametric based methods extend intensity-based approaches by modeling the intensity as a signal.

Regardless which of the three methods of detecting interest points are employed, detected points are expected to embody certain attributes. T. Tuytelaars and K. Mikolajczyk found that regardless of implementation domain and application, in general, detectors based on the principals of the following properties should target interest points good for tracking applications:

Repeatability: Given two images of the same object or scene, taken under different viewing conditions, high repeatability indicates that a high percentage of the interest points that are visible in both images are detected in both images. Conversely, low repeatability means that only a small number of the interest points that are detected in one image and are visible in the other image are not detected in the other image.

Distinctiveness: The uniqueness of an interest point embodies how well it can be matched. The intensity patterns underlying the detected interest points should show a lot of variance, such that interest points can be distinguished.

Locality: The interest points should be local, so as to limit the risk of a interest points including an occluded part and/or parts corresponding to different objects or surfaces, and to allow simple approximations of the geometric and photometric deformations between two images taken under different viewing conditions.

Quantity: The number of detected interest points should be sufficiently large, such that a reasonable number of points are detected even on small objects.

Accuracy: The detected interest points should be accurately localized, both in image location, as with respect to scale and possibly shape.

Efficiency: Preferably, the detection of interest points in a new image should take just fractions of a second, to allow for time-critical applications.

Robustness: In case of relatively small deformations, it often suffices to make sure the interest points detection methods are not too sensitive, i.e. the accuracy of the detection may go down a bit, but not drastically so. Robustness is defined in this case with respect to image noise, discretization effects, compression artifacts, blur, etc., as well as geometric and photometric deviations from the mathematical model used to obtain invariant interest points.

The performance metrics listed below are designed to quantify an interest point detector's capability to seek out and capture the desirable attributes defined above. Previous research [10], [11], [12], [13] employed these metrics with some minor differences, specific emphasis, and some omissions.

Recall: Calculated by dividing the number of correct matches by the number of total correspondence, this score measures a descriptor's ability to produce correct correspondences.

Min-recall: Calculated by imposing a lower bound on recall, since perfect matching is not useful if too few matches are made to further perform calculations. For example, a minimum of four points are required for a homography and a minimum of seven points are required for a fundamental matrix. The concept and calculations of a homography and a fundamental matrix are explained later in this chapter.

1-Precision: Calculated by dividing the number of incorrect matches by the number of correct matches plus the number of incorrect matches, this score measures the inverse of a descriptor's exactness or fidelity. When considered against the independent recall measurement, a comparable performance curve is formed. See [10] and [11] for more details.

Receiver Operating Characteristics (ROC): Calculated as the detection rate verses the false positive rate, the ROC curve provides a measure of true positives to false positives.

Repeatability: Calculated for a pair of images as the ratio between the number of region-to-region correspondence and the smaller of the number of regions in the pair of images, repeatability describes how well the regions of interest are similar from image to image. Two regions are deemed to correspond if the overlap error, defined as the error in the image area covered by the regions, is sufficiently small.

Accuracy: Calculated as a relative ranking among descriptors, accuracy is determined by a function of an overlap error; if the overlap threshold is relaxed more regions correspond and repeatability goes up. If a descriptor improves in recall as a result, the descriptor is ranked as "less pixel-wise accurate" than the others are due to the relaxation of the threshold.

Distinctiveness: Calculated by generating eigenvalues from the PCA of the descriptors normalized by their variance, this metric demonstrates the relative discrimination power of a descriptor.

In the following section, we describe the implementation approaches of three representatives of the most common interest point detector types.

a. Maximally Stable Extremal Regions (MSER)

First proposed by Matas et al. [22], MSER finds interest points defined by image elements coined extremal regions. Extremal regions are unique in that they are closed under continuous, projective transformations of image coordinates and under monotonic transformation of image intensities. Figure 2 describes the formal concept of operation for MSER.

Image I is a mapping $I : \mathcal{D} \subset \mathbb{Z}^2 \rightarrow \mathcal{S}$. Extremal regions are well defined on images if:

1. $\mathcal{S}$ is totally ordered, i.e. reflexive, antisymmetric and transitive binary relation $\leq$ exists. In this paper only $S = \{0, 1, \ldots, 255\}$ is considered, but extremal regions can be defined on e.g. real-valued images ($\mathcal{S} = R$).
2. An adjacency (neighbourhood) relation $A \subset \mathcal{D} \times \mathcal{D}$ is defined. In this paper 4-neighbourhoods are used, i.e. $p, q \in \mathcal{D}$ are adjacent (pAq) iff $\sum_{i=1}^{d} |p_i - q_i| \leq 1$.

Region $\mathcal{Q}$ is a contiguous subset of $\mathcal{D}$, i.e. for each $p, q \in \mathcal{Q}$ there is a sequence $p, a_1, a_2, \ldots, a_n, q$ and $pAa_1, a_iAa_{i+1}, a_nAq$.

(Outer) Region Boundary $\partial\mathcal{Q} = \{q \in \mathcal{D} \setminus \mathcal{Q} : \exists p \in \mathcal{Q} : qAp\}$, i.e. the boundary $\partial\mathcal{Q}$ of $\mathcal{Q}$ is the set of pixels being adjacent to at least one pixel of $\mathcal{Q}$ but not belonging to $\mathcal{Q}$.

Extremal Region $\mathcal{Q} \subset D$ is a region such that for all $p \in \mathcal{Q}, q \in \partial\mathcal{Q} : I(p) > I(q)$ (maximum intensity region) or $I(p) < I(q)$ (minimum intensity region).

Maximally Stable Extremal Region (MSER). Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{i-1}, \mathcal{Q}_i, \ldots$ be a sequence of nested extremal regions, i.e. $\mathcal{Q}_i \subset \mathcal{Q}_{i+1}$. Extremal region $\mathcal{Q}_{i^*}$ is maximally stable iff $q(i) = |\mathcal{Q}_{i+\Delta} \setminus \mathcal{Q}_{i-\Delta}|/|\mathcal{Q}_i|$ has a local minimum at i^* ($|.|$ denotes cardinality). $\Delta \in \mathcal{S}$ is a parameter of the method.

Table 1: **Definitions** used in Section 2

Figure 2. MSER interest point detector elements. [From [22]] D is a set of continuously varied grayscale threshold images; S is the set of extremal regions.

Essentially, if we perform grayscale thresholding of an image, we can segment all pixels into those that lie below the threshold and those that lie above the threshold. As we continuously vary the threshold from one extreme (all above or all below) to the other, a sequence of segmentation is produced. The connected components of the sequence are extremal in the sense that they extend from one extreme threshold to the other. This set is defined as extremal regions. MSER demonstrated the best overall performance capability in achieving viewpoint invariance in the recent performance evaluation of Mikolajczyk et al. [11].

b. Harris Corner Detector

An interest point detector is primarily an attribute targeting mechanism. The detector finds scene attributes based on domain-dependent

properties that will lead to a reliable descriptor. Although there are many implementations, the Harris corner detector [34], the Difference of Gaussian detector [35], and the Lucas-Kanade detector [36] stand out as class representatives [13]. From the binary corner class of detectors, the Harris corner detector [34] is one of the most widely used detectors. This detector operates by first finding a covariance matrix through a Gaussian filter convolution of the image. Then a second-moment matrix is generated through the 2nd Gaussian derivatives of the covariance matrix. The eigenvalues of the second-moment matrix represent the strength of the gradient in image intensity parallel and perpendicular to the direction of the greatest change. Two large eigenvalues corresponding to a strong change in any direction define a corner. The Harris algorithm considers points with local maxima eigenvalues as corner points.

c. Difference of Gaussian (DoG) Detector

To detect appropriate points in a scene, Lowe [35] implemented a Gaussian-smoothed pyramid of images with increasing scale. The DoG maxima identify the points of interest. Another detector also worthy of note was developed by Lucas and Kanade [36], as a gradient decent method that is used to iteratively align image intensity patches.

3. Interest Point Descriptors

After the detection phase, where we identify interest points in an image, we need to be able to uniquely describe each point. As we stated earlier in this chapter, there are many different forms of interest point descriptors. Most descriptors contain information that describes the interest point orientation, strength, and scale allowing for transformation invariance while still capturing salient region information. The most basic descriptor is comprised only of the vector of image pixels [14] in the region of interest. While this descriptor allows for a simple correlation of interest points through a computation of a similarity score, the lack of non-translational transformation invariance limits its suitability

in most applications. On the other hand, transformation-invariant descriptors are typically high-dimensional vectors requiring significant computational complexity. Choosing the appropriate descriptor for an application is a matter of striking a fine balance between correspondence performance and processing efficiency. Below we describe distribution-based, spatial frequency-based, differential-based and gradient moment descriptors.

a. Scale Invariant Feature Transform (SIFT)

The SIFT algorithm was proposed by Lowe [20], [31] as a method of extracting and describing interest points (which Lowe calls key-points). This description process is robustly invariant to scale transforms, but is also invariant to other common image transformations and deformations such as image rotation, illumination changes and blur. The SIFT algorithm has four major stages: Scale-space extrema detection, Keypoint localization, Orientation assignment and Keypoint descriptor.

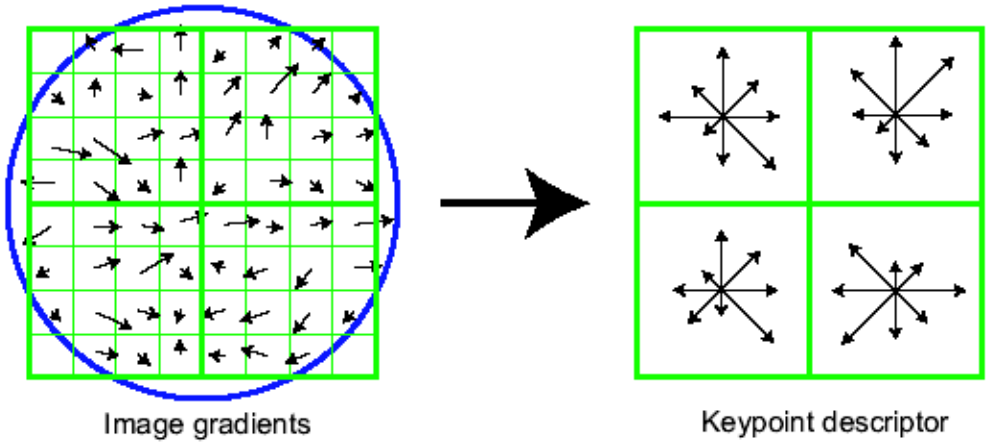


Figure 3. SIFT descriptor elements. [From [20]]

1. Scale-space Extrema Detection: In this stage of processing, SIFT performs a search over all scales and image locations in the image. The goal is to find each scale-space extrema point through the

Difference of Gaussians (DOG) function $D(x,y,\theta)$, which can be computed from the difference of two nearby scaled images separated by a multiplicative factor k:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \; = L(x,y,k\sigma) - L(x,y,\sigma)$$

Here, $L(x,y,\sigma)$ is the scale space of an image, built by convolving the image $I(x,y)$ with the Gaussian kernel $G(x,y,\sigma)$, Bingrong et al. [37].

2. Keypoint Localization: At each candidate location, a detailed model is fit to determine the exact sub-pixel location and scale. Interest points (a.k.a. Keypoints) are selected based on measures of their stability.

3. Orientation Assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been normalized relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations. One or more orientations are assigned to each key-point based on local image gradients. For each image sample $L(x,y)$ at this scale, the gradient magnitude $m(x,y)$ and orientation $\theta(x,y)$ is computed using pixel differences :

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}((L(x+1y) - L(x-1,y))/(L(x,y+1) - L(x,y-1))) \text{ [37]}$$

4. Keypoint Descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. Typical keypoint descriptors use 16 orientation histograms aligned in a grid [37]. Each SIFT histogram has eight orientation bins created over a 4X4 support window. The resulting interest point vectors are 128 elements with a total support window of 16×16 pixels. In SIFT, the number of generated interest points depends on image size and content, as well as algorithm parameters. An image of size 500x500 pixels will yield about 2000 stable features. For a more detailed description, see [20], [31].

b. Principal Component Analysis (PCA)-SIFT

PCA [38] is a common vector space transform technique that has been applied to a wide variety of computer vision applications in order to perform dimensionality reduction. Defined mathematically as an orthogonal linear transform, PCA transforms vector data into a new lower-dimensional coordinate system, preserving the variance in the original data as much as possible. Ke and Sukthankar's [39] PCA-SIFT combines this idea with SIFT to produce a lower-dimensional feature descriptor with similar characteristics. It accepts the same input as the standard SIFT descriptor, specifically the sub-pixel location, scale and dominant orientations of the interest point. In PCA-SIFT, a 41 × 41 patch is extracted at the given scale, and rotated to align its dominant orientation to a canonical direction. The PCA-SIFT process involves pre-computing an eigenspace to express the gradient images of local patches. With each candidate patch, the local image gradient is computed and the gradient image vector is projected onto the eigenspace to derive a compact feature vector. This feature vector is significantly smaller than the standard SIFT interest point vector, and can be used with the same matching algorithms. For a more details, see [39].

c. Speeded Up Robust Features (SURF)

The most recently developed descriptor in our inventory is SURF [40]. SURF operates in a similar fashion to the SIFT descriptor, except that the published version is closely coupled with an interest point detector that is based upon generating a Hessian matrix approximation of integral images. This approach drastically reduces computational time in that rather than using two different calculations for selecting the location and the scale of an interest point, SURF relies on the determinant of the Hessian matrix for both. Given a point x = (x, y) in an image I, the Hessian matrix $H(x,\sigma)$ in x at scale σ is defined as follows

$$H(x,\sigma) = \begin{pmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{pmatrix}$$

Here, $L_{xx}(x,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point x, and similarly for $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$.

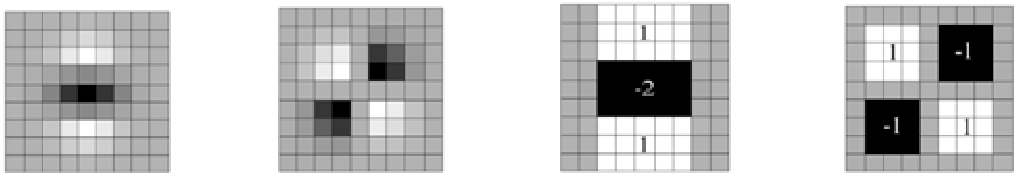


Figure 4. Left to right: the Gaussian second order partial derivatives in y-direction and xy-direction, and approximations thereof using box filters. The grey regions are equal to zero. [from [40]]

To produce a SURF descriptor, Haar wavelet responses, centered on the Hessian detected interest points, are generated in the x and y directions. The characteristic scale is determined as part of the SURF detection and is used to determine the Haar sampling step. The Haar responses are assigned Gaussian weights and are then used to determine a characteristic orientation. Once the orientation is assigned, 4 X 4 square regions are defined at each interest point location and then further divided into regular 4 X 4 sub-regions. The 64-dimensional descriptor vector is generated by summing over Haar responses for the sub regions. The vector includes dimensions that account for intensity changes by recording the absolute response values in the x and y directions. If the problem domain requires the additional invariance provided by a 128-dimension SURF descriptor, the sub regions can be further sub divided into 4 X 4 sub-sub regions.

d. Gradient Location and Orientation Histogram (GLOH)

Mikolajczyk and Schmid introduced GLOH in [11] as an extension of the SIFT algorithm. The GLOH description process operates on detected interest points by first computing the SIFT descriptor for a log-polar grid with 3 bins in a radial direction and 8 bins in an angular direction for 17 location bins. The gradient orientations are then quantities into 16 bins. The result is 272-bin histogram. PCA is then used to reduce the dimensionality of the descriptor to 128 dimensions.

e. Shape Context

The shape context descriptor is described by S. Belongie et al. [41], and is also modeled after the SIFT descriptor except it functions on 3-D histograms of edge point locations and orientations. At each interest point, edges are extracted by a Canny edge [42] detector, and the interest point region is quantized into nine log-polar bins. These bins are then divided into radii of 6, 11 and 15, and the orientation is quantized into horizontal, vertical and two diagonal bins, resulting in a 36-dimensional descriptor.

f. Gradient Moments

Gradient moments were first introduced by Van Gool, et al [43] and are designed to characterize the intensity patterns of image regions. These intensity moments are largely invariant to affine transformations for simple shapes and do not require the computationally expensive extraction process that the distribution-based methods describe above do. In this application, moments are computed for the derivatives of a grayscale image patches as follows:

$$M_{pq}^{a} = \frac{1}{xy} \sum_{x,y} x^{p} y^{q} [I_{d}(x, y)]^{a}$$

Here, $p + q$ is the order, a is the degree, and I_d is the image gradient in direction d. The second order, second-degree moment gradients are calculated in the x and y directions. The result is a 20 dimensional feature vector.

g. Normalized Cross Correlation Template Matching

Normalized Cross Correlation Template Matching (NCCTM) [76], is perhaps one of the most basic spatial frequency based descriptor methods. A NCCTM descriptor is created either through the spatial domain or through a transform domain. For the spatial domain, the region surrounding an interest point is sampled uniformly, in say a 9 X 9 pixel pattern. The spatial-domain method of cross correlation has two drawbacks; the convolution of a NCCTM template with a bright spot may produce a higher correlation than with a matching patch and NCCTM templates are not invariant to lighting changes. For this reason, NCCTM achieves better performance through a transform of an image such as through Fast Fourier Transform [76].

h. Steerable Filters and Differential Invariants

. Freeman and Adelson's Steerable filters [44] are computed by taking the derivative of up to the forth order through convolution with Gaussian filters with σ=6.7 for an image patch of size 41. This results in a feature vector dimension of 14. Koenderink and van Doorn's differential invariants [45] are computed in the same manner except only up to 3rd order for a feature vector dimension of 8.

i. Complex Filters

In Schaffalitzky and Zisserman's approach [46], a complex filter kernel is employed to generate a feature vector description. The kernels are implemented as a filter containing a unit disk of radius 1. The filter convolution is performed in a 41X41 patch about the interest point from the following equation:

$$K_{mn}(x,y) = (x+iy)^m (x-iy)^n G(x,y)$$

4. Interest Point Matching

We have described the basic tenants of finding and describing interest points in images, and are now interested in matching them. For humans, a matching interest point can be found in two images relatively accurately, such as pinpointing the corner of a window as projected in two different images. For a computer there needs to be an algorithmic method for accomplishing the match since a computational machine does not contain the conceptual knowledge of a corner or a window. Successfully finding a correspondence of points between images is central in the design of vision-based tracking algorithms. Correspondence computation comes from the knowledge that the points in different images represent projections of the same point onto image planes. Autonomously establishing corresponding points between any two images begins by finding good interest points in the first image. Ideally, an algorithm is able to discern objects or object attributes in an image without any prior information about them, and then proceed to track them between progressive frames by recognition regardless of inherent transformations in pose and illumination. Invariance properties like these have become the standard for which interest point extraction, description and matching algorithms are measured [10], [11], [12], [13]. The primary goal of any such extraction algorithm is to maximize image transformation invariance using techniques described earlier in this chapter. Optimizing the employment of each algorithm based on its particular invariance for a given environment is the focus of this thesis.

For establishing corresponding points between any two images, there are two general approaches. In some vision tracking applications such as the process of generating 3D structure from images, described as Structure from Motion (SFM) in [47] and the real-time application of SFM also known as SLAM proposed by [48], the by-frame processing latency is low enough to support a high data rate video stream. This high data rate results in a short transfer

distance between any two images. If the application only requires a superficial knowledge of each interest point, then in this case the tracking problem reduces to producing chains of correspondence across frames. As mentioned in opening of this chapter, this is known as a narrow- or short-baseline correspondence problem and is addressed by several tracking methods [18], [49], [50]. Short-baseline correspondence is not adequate in cases where only a sparse set of widely separated images are available to process. Another more typical case is that we desire to preserve the interest points for tracking through occlusion or for long-term retrieval and object identification and matching. In these, wide-baseline situations, interest points are found in the first frame and a unique descriptor is generated for each point [51]. We then repeat the same processing on the second image, and establish correlations between the set of features from the first to the second image and consider these corresponding points if they match within a certain geometric distance threshold.

For wide-baseline tracking, there are a number of ways to find the "distance" between two potential matching feature vectors. Four common methods are Nearest Neighbor with Distance Ratio (NNDR), Normalized Cross Correlation (NCC), Sum-of-Squared-Difference (SSD) and specific to Kanade-Lucas-Tomasi (KLT) tracking approach is the KLT matching [13]. In NNDR, the Euclidean distance between normalized interest point descriptors is calculated and compared to neighboring values to determine a match [35]. NCC is a threshold-based matching algorithm where a correlation coefficient between interest point descriptors is calculated and the coefficients that are above a certain threshold are considered matches. SSD is also a threshold-based matching algorithm and is calculated by summing over the difference of interest point descriptors. If the generated distance is below a certain threshold, the pair is considered a match. A gradient descent method is used to the align image patch descriptors in successive images [49].

5. Performance Evaluation

Interest point tracking is a complex problem. External variables and constraints inherent in most application environments complicate the process. For example, intuitively interest points can move out of the field of view, an image can have repetitive patterns, interest points can blur as a result of camera motion, tracked points may be occluded in one of the images and tracked points may not be revisited for many frames. Recent research efforts [10], [11], [12], [13] have essentially sought to quantify the performance of interest point detectors and descriptors given data sets which specifically test the invariant, or covariant, properties of each implementation. Conversely, this thesis seeks to evaluate the performance of interest point descriptors within different scene classes. In the following section, we discuss camera transformation ground truth determination and data set development utilized in previous work. This will provide the reader with a solid background for considering our ground truth determination presented in the next chapter.

a. Ground Truth

One point of departure among the previous interest point evaluation research methods is the specific method for determining the ground truth of the camera transformation. While [10] and [11] have focused on employing a planar homography between two images to determine ground truth, [13] opted for finding a fundamental matrix and [12] a trifocal tensor and point transfer property to find a non-planar homography. The remainder of this section details these concepts and processes. For additional information, refer to [52].

A 2-D image homography is a projective transformation mapping of points from one image plane to another image plane. 2-D homographies have 8 degrees of freedom with nine entries formatted in an H matrix. Scale is the unrecoverable ninth entry in the matrix. A homography requires a minimum of four pairs of corresponding points, in a feature tracking application we use interest points. If any three points are collinear, the result will not find a unique

solution. To find a homography, we solve for an *H* that satisfies $X_E = H * X_P$, where X_E and X_P are an ordered set of corresponding image point pairs.

A fundamental matrix is similar to a homography in that it is a projective transformation mapping of points form one image to another, however a fundamental matrix is capable of modeling non-planar 3-D transformations. A fundamental matrix is based upon a geometric relation of the point transformations. This relationship is referred to as epipolar geometry. To find a fundamental matrix, we solve for an *F* that satisfies $x_1^T F x_2 = 0$, where x_1 and x_2 are homogeneous corresponding image point pairs. For additional information, see [72].

To find ground truth with Fraundorfer and Bischof's [12] trifocal tensor and point transfer property, an image sequence is defined as $I_1 - I_n$, and a geometrically correct mapping is found for every detected location in I_1 to the other ***n*** images in the sequence. The key to this method is that with a set of three images containing portions of the same scene, say I_1, I_2 and I_3, it is possible to calculate the point x_3 in I_3 where x_1 and x_2 are corresponding image point pairs from I_1, I_2. The trifocal tensor and the point pair are then used to compute a non-planar mapping that projects the point from image I_1 to the target image.

b. Data Sets

To properly evaluate the performance of interest point descriptors, experiment data sets of images have to be carefully designed to demonstrate the camera transformations to be tested. Mikolajczyk, et al. [10], [11] employed a data set with two scene types. One scene type contains homogeneous regions with distinctive edge boundaries (e.g. graffiti, buildings), and the other contains repeated textures of different forms. The data set includes viewpoint changes, scale changes, image blur, JPEG compression and illumination. In the cases of

viewpoint change, scale change and blur, the same change in imaging conditions is applied to the two different scene types. In the viewpoint change test, the camera varies from a front to parallel view to one with significant foreshortening at approximately 60 degrees to the camera. The scale change and blur sequences are acquired by varying the camera zoom and focus respectively.

Fraundorfer and Bischof [12] acquired two image data sets each with 19 images taken from viewpoints varying from 0º to 90º. The first image set is piece-wise planar and shows two geometric boxes posed on a turntable. The second image set captures a part of a room.

Klippenstein and Zhang [13] produced two image data sets for their research by manually driving a robot with a digital camera through different scale environments. The large-scale image set was generated on a building floor that is typical of an office building. The robot was operated for 30 meters with an image captured at every 150mm translation or 5º rotation. The small-scale image set was acquired in a research lab with an image captured every 100mm or 5º rotation.

D. TRACKING APPLICATIONS

As mentioned in Chapter I, numerous tracking applications utilize computer vision as the primary sensor. This section provides some examples of tracking applications.

1. Simultaneous Localization and Mapping (SLAM)

Simultaneous Localization and Mapping (SLAM) was originally presented by Hugh Durrant-Whyte and John J. Leonard [53]. The domain of SLAM is concerned with real time Structure from Motion (SFM), which is essentially the problem of building a map of an unknown environment by an autonomous mobile robot while simultaneously navigating the environment using the map. SLAM consists of several individual processes: landmark discovery and extraction, data

association, state estimation, state update and landmark update. There are many ways to solve each of the SLAM components. Below we describe some vision-based SLAM systems.

a. FastSLAM Based MonoSLAM

Eade and Drummond [54] have developed a SLAM algorithm based on Montemerlo, et al's FastSLAM [55], using a single camera as a sensor. They demonstrated a camera traversing a circular pattern in a three-dimensional area, mapping multiple features with a successful closing of the loop. FastSLAM is an algorithm developed for traditional range based sensor SLAM that uses particle filters. A particle represents a probabilistically weighted pose of the robot, and each is composed of a historic path estimate and a covariance matrix coupled with a set of estimators of individual interest point locations. New readings update the particles and cull those that are probabilistically unlikely. The demonstration was proof of the usefulness of traditional SLAM algorithms to vision based SLAM.

b. MonoSLAM

Andrew Davidson's MonoSLAM [56] is an approach that uses a monocular camera to conduct SLAM by way of a sparse map, which means the actual map itself is a point cloud of the 3-D Euclidean location of the tracked interest points. Essentially the system seeks to track a sparse number of image patches and refine their positions through camera motion in 3-D grid map. Using quaternion patches to define the interest points, the algorithm is able to both store and reacquire patches when they come back into view. The focus of the algorithm is a real-time, monocular vision based, SLAM algorithm. An added feature of MonoSLAM over a normal occupancy grid approach is that their features correspond to a covariance matrix, which defines the probability of the true location of the feature.

E. IMAGE CLASSIFICATION

Image classification is a research discipline within computer vision that has seen a recent explosion of research efforts [57], [58], [59], [60], [61], [62], [63], [64]. Early image content classification experiments centered on the task of classifying objects or foreground and background components of visual scenes into relevant, semantic categories. Drawing on the previous work of Campbell, et al [57], Israël, et al [65] developed a process for dividing an image into a group of adjacent texture pixels, each as a semantic image element. These elements, or patches, are categorized as building, grass, crowd, road, sand, skin, sky, tree, or water. The patches comprise an individual image feature vector that is processed to provide an overall scene classification of interior, city/street, forest, agriculture/countryside, desert, sea, portrait or crowds.

More closely related to our work, Rasiwasia and Vasconcelos [58] described a process where the task of automatic scene classification is accomplished by first defining a method of image representation and then employing a weak-supervision machine learning classifier for determining the appropriate category. The image description is composed of a bag of low dimensional localized descriptors such as spatial frequency or SIFT-like descriptors. The descriptor coefficients are grouped into semantically relevant themes called visterms. The visterms are utilized to classify scenes into learned classifications through Support Vector Machines (SVMs) [66]. The benefit of this research is a well-performing image theme classification method based upon a low-dimensional feature vector.

In this chapter, we have explored the process of human motion perception, visual interest point extraction and tracking, and finally, scene classification. This discussion will provide the reader a solid background of knowledge for the next chapter, experiment design.

III. EXPERIMENT DESIGN

As stated in Chapter II, it is our intent to present a unique approach to evaluating interest point extraction and correspondence performance by looking at specific scene classes. To that end, we have tested and compared an inventory of extractors. We have focused on a test methodology that achieves highly reproducible conditions and a data set collection methodology that generates scene image sets with consistent transformations across each scene. In particular, we have employed publicly available software and libraries wherever possible in our testing. Our testing platform includes a PC with dual Intel Xeon QuadCore CPUs and 16GB RAM, with Redhat Linux and a Windows Vista Operating Systems with MS Visual Studio® 2005 and MATLAB®. Additional software libraries include the SURF® extraction library [40], Robert Hess's SIFT implementation [67], Krystian Mikolajczyk's feature detection [10] and description [11] library, as well as Intel's OpenCV [68], and GNU GSL [69]. The following sections describe our experiment setup.

A. IMAGE DATA SET

A few well-known and well-tested image data sets exist and have been utilized in previous research efforts [10], [11], [12], [13]; however, these data sets were generated to evaluate the overall performance of feature extractors given the viewpoint invariance properties described in Chapter II. Since we seek to evaluate the specific performance of each extraction algorithm and technique for a specific scene class, these image data sets are insufficient. We have developed an image data set containing a sequence of consistent camera transformations captured from different scene classifications. The following sections provide specific data set details.

1. Scene Types

For our research, the scene categories found in [58] inspired us: office, living room, bedroom, kitchen, store, industrial, tall building, inside city, highway, coast, open country, mountain, forest and suburb scene classifications. Since our goal is not scene classification, but instead is to determine the best feature detector/descriptor to use for each scene category, we broke up the categories slightly differently. As noticed by previous research in semantic scene classification [58], indoor scenes can vary greatly in visible attributes and require sub-categories. Based on the presence of an indoor scene's displayed texture pattern and color change frequency variance and intensity (i.e. the frequency of variance of observed light wavelength,) we formed the sub-categories of *interior dense* and *interior sparse*. We additionally added *urban short building* and *high desert* scenes. For our images depicting an *indoor dense* scene, we captured a sequence of images in the Grand Ballroom of the Del Monte Hotel in Monterey, CA, centered on a water fountain. For an *indoor sparse* scene, we used a wall in a typical foyer at the Naval Postgraduate School (NPS). All outdoor scene image sequences were captured on the NPS campus and in the area surrounding Monterey, CA and Kernville, CA.

2. Image Acquisition

In support of the goal of evaluating extractor performance based on scene type, our experiments required an image sequence that includes camera transformations (scale, translation and rotation) that can also be easily recreated in multiple scene environments. Since the metric range scales encountered in indoor and outdoor scenes vary widely, we chose to design a pattern that could apply to both indoor and outdoor environments by merely adjusting the scale of the pattern.

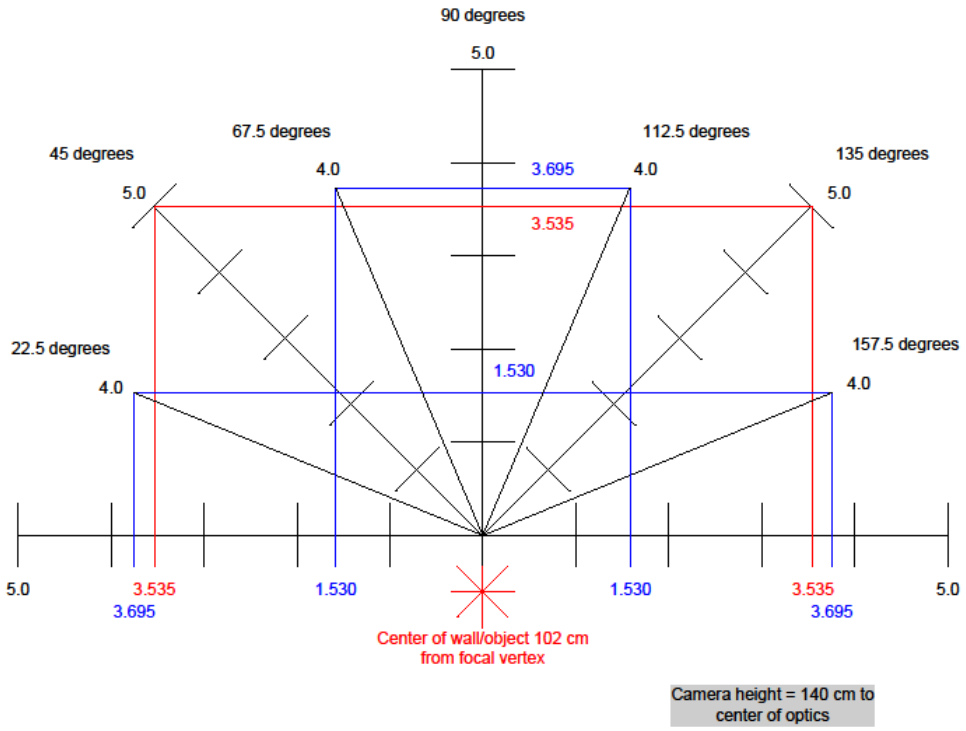


Figure 5. Image capture pattern.

We started by assuming that our sequence would begin with a reference image whose captured scene would mostly be contained within all the other images. This is easily accomplished for relatively shallow scenes such as walls. Scenes with great depth such as landscapes have less overlap between the reference image and other images. The advantages of non-planar, “deeper” scenes are more complex, non-affine transformations of scene content, which has higher demands on the interest point descriptors.

Our capture pattern is centered on and symmetric about an object or aim point in the center of the scene we desire to capture. For indoor scenes, our center point is fixed at 102 cm from the focal vertex, for outdoor scenes, it is 1 meter. The central object or benchmark gives us an aim point for each frame.

To demonstrate the camera transformations we wish to include in our experiments, we have opted for a radial pattern with seven sequence paths on which we situate our camera and capture our images. As illustrated in Figure 5, the paths that are at +/- 45 degrees and the path that is perpendicular to the plane at the center of the scene are divided into five equidistant unit positions. The remaining paths that are oriented at 22.5 degrees and at 67.5 degrees contain only one capture point each, located at the end of a four-unit path. Hence, there are seven measurement points for four-unit distances, arranged circularly around the focal vertex. For indoor sequences, we chose one unit of measurements on the pattern to equal one meter. For outdoor sequences, we centered our pattern five meters from the wall/object and we let one unit of measurement equal four meters.

To capture each scene in a set of images, we employed a 7.1-megapixel Canon Powershot Elph SD1000 with optical image stabilization. The camera was mounted on a tripod and triggered manually. We began each experiment run by first establishing the sequence pattern in the environment to be captured. The camera and tripod were then aligned to the location of the first position of the capture sequence and the viewfinder was used to aim the camera optics directly at the center of the central object or benchmark of the scene. An image was captured at each location specified by the sequence pattern in this fashion.

Our pattern accounts for scale, translation and out-of-plane rotation transformations. To demonstrate in-plane rotation transformations within the data set, we formed synthetic rotations through image editing software. New images were created by rotating the captured sequence shots to 035º, 160º, 200º, and 325º from the initial capture position as shown in the upper left and upper right sample images in Figure 6. The rotation operation preserves the image resolution and scale.

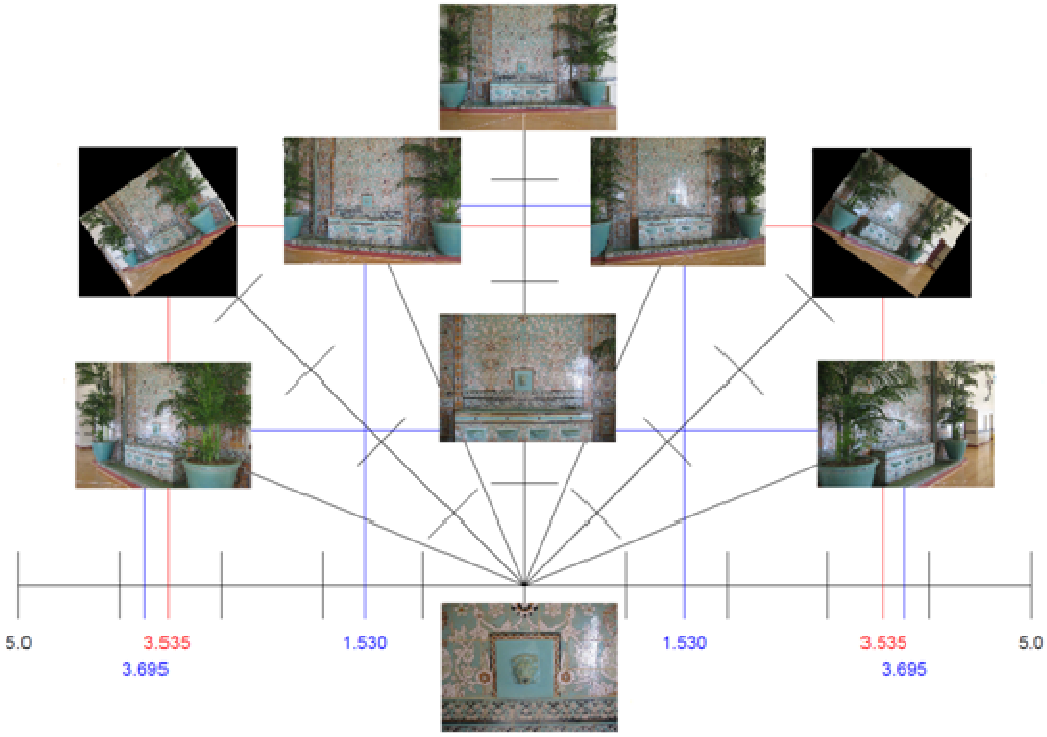


Figure 6. Sample data set images. The 045.0 image was rotated 325º, and the 137.0 image was rotated 035º.

B. IMAGE PROCESSING METHODOLOGY

Since we are evaluating how to best employ currently existing extractors given our data set, we first obtained implementations of each extractor that we desired to include in our experiments from the internet (see [67], [68], [69], [70].) We drew upon the sample source code that was available with the SURF and the Hess SIFT extractors to create an image-processing test framework. The following sections detail the logic behind the framework design and implementation.

1. Interest Point Detection

As noted in Chapter II, interest point extraction involves two steps, point detection and point description. The first functional section of our image-processing framework includes image pre-processing and interest point

detection. Before we can find interest points within a scene, the captured image requires some minor preparation and preprocessing operations such as size normalization and (for our framework) converting from color to grey scale.

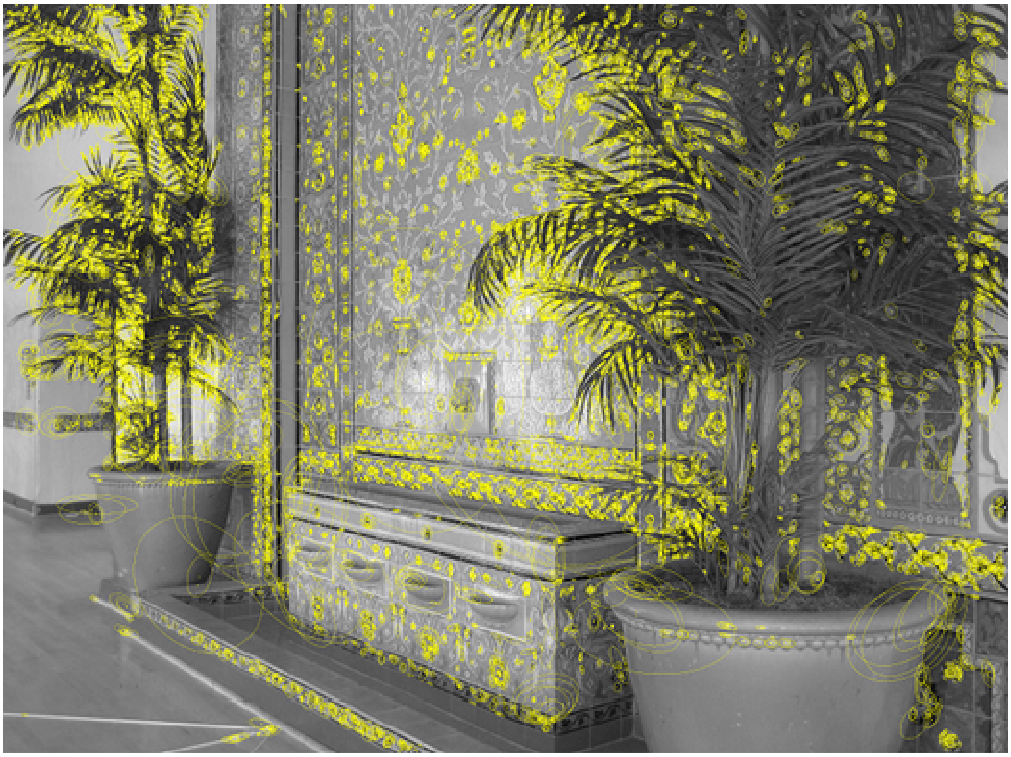


Figure 7. Detected interest points plotted on a data set image.

Mikolajczyk, et al. [10], [11] found that in the case of evaluating extractor performance, the specific detector employed was less of a performance factor than the descriptor algorithm employed. That said, the MSER detector exhibited the best overall performance results. Therefore, we decided to employ it for each descriptor tested except SURF and Hess SIFT in our experiments. The SURF descriptor is tightly coupled to a Hessian-based point detector as discussed in Chapter II. To provide a point of comparison between performance based upon MSER generated interest points and other detection means, we employed the DoG detector as described by Lowe [20] for the Hess SIFT descriptor. Using the author recommended detector settings of ellipse-style regions and a scale size of two, MSER interest points are detected within all the images of the capture

sequence and then stored as interest point detection files for future processing. As shown in Figure 7, the MSER detection process generates regions, centered on a point of interest. (For more detail on the MSER detector, see Chapter II.)

2. Interest Point Description

After detecting the candidate regions within all the images and after saving a listing of those interesting points, we need to generate the interest point descriptors. With the corresponding MSER detection file for each image, we generated appropriate feature vectors for each description method. The descriptor algorithms are executed with the respective author-recommended options and properties. The descriptors are then stored in files for future processing.

Since our target application is vision-based SLAM, the most relevant descriptors are based on viewpoint invariance and provide the greatest potential for long-term recall. We have selected 10 descriptors: two versions of SIFT, PCA-SIFT, SURF, GLOH, Gradient Moments, Shape Context, Cross Correlation, Steerable Filters, Differential Invariants, and Complex Filters.

3. Interest Point Correspondence

After we have generated the interest points for each image, our next task is to find matching interest points between two images. For this research, we need to compare each image against the reference image. For all scenes, the image that was captured in the center and closest to the scene center object or benchmark was designated as reference image. For the *urban short building* scenes, we also conduct additional experiment runs using the image captured in the center furthest way from the scene center or benchmark for the reference image. For matching efficiency, we use the reference image descriptor vectors to build a KD tree. Beginning with the dimension that has the greatest variance, the image descriptor files are loaded and then the nearest points for each vector are searched for in the KD tree. Inspired by Hess [67], the comparison method

will find the two nearest neighbors of the input vector within the tree by using a Best Bin First (BBF), N-nearest neighbor search. We compute the vector distance of the two closest match neighbors by finding the sum of the squared difference (SSD) of each individual dimension. If the vector distance of the first nearest neighbor is at least twice as "far" as the vector distance of the second nearest neighbor, then the first nearest neighbor is considered a potential match. This approach has yielded good performance in empirical evaluations of previous research [10], [11], [12], [13]. We attempt to store each potential match in a map data structure and perform a reverse-lookup. That is, if the map indicates that the matched nearest neighbor is already contained within the map, that is if the interest point has already been matched to a different point, then the match with the least vector distance is determined to be the best match and the other match is removed from the map. The result is a map containing point correspondences for each reference-image to image pair.

C. CAMERA MOTION DETERMINATION

The next functional step in our experiment is to determine the 3-D motion of the camera for each reference-image to image pair. As with previous performance evaluations [10], [11], [12], [13], in this research, a determination of camera motion will serve as ground truth for our experiments and will be used to evaluate the correctness of the matches produced by each extraction algorithm. The objective of this process is to find outliers, meaning incorrect correspondences, not to validate the precise position of an interest point. The assumption is that the transformation of correct correspondences can be successfully modeled with a fundamental matrix and that the fundamental matrix can be computed with standard methods. This assumption allows us to use the fundamental matrix as a model of interest point movement between two images.

Additionally, since our images are significantly larger in number (seven scenes of 99 images each) and size (7.1 mega-pixels) than images used in previous experiments (e.g., [11]), we found it inefficient to employ a hand-based

method of determining ground truth. Because of the size of our images, the MSER detector typically generated many interest points. For example, an image in our high-desert stump scene generated 18,066 interest points. Instead of altering the detection threshold to reduce the number of interest points and possibly change the results of our experiment, we developed a method that utilizes all the detected points. Additionally, since we expect to see large, out-of-plane transformations in our images, we use an 8-point Random Sample Consensus (RANSAC) [71] to generate a fundamental matrix that describes the ground truth camera motion. This way we can utilize all the detected points in our evaluation. Each RANSAC generated Fundamental matrix is visually verified to ensure it accurately describes the camera transformations. We describe our RANSAC approach in the next section.

1. RANSAC Method

RANSAC is a robust parameter estimation algorithm introduced by [71] that iteratively fits a mathematical model to experiment data. In our application, for each reference-image to image pairing, our data is comprised of the point correspondence map and our model is a transformation matrix in space producing each match. RANSAC iteration detailed in Figure 8 develops a convergence on a fundamental matrix that provides for the translation of one camera position to another. As a result, the RANSAC algorithm will produce inliers that fit the model and outliers that do not fit the model within a specified tolerance of 8-pixels in re-projection. This approach is not consistent with the 1.5 to 3-pixel tolerances found in the work presented in [10], [11], [12], [13]. However since we are not implementing an interest point region overlap method of determining an overlap error as in [10], [11], [12], [13], and because our images are much larger, a 1.5 to 3-pixel threshold is not sufficient to account for interest points with larger regions. To determine an appropriate error threshold, we projected interest points from the reference image forward and from the subject image backward by transforming each set of descriptor-matched

corresponding points with the fundamental matrix. We then produced a plot (such as the one shown in figure 11) of distances of normalized transformed point to the corresponding epipolar line (See [72] for additional information on epipolar geometry) and determined that a threshold of 8-pixels was conservative and appropriate for our data set.

Determine:

***n* – the smallest number of points required**

***k* – the number of iterations required**

***t* – the threshold used to identify a point that fits well**

***d* – the number of nearby points required to assert a line fits well**

Until *k* iterations have occurred:

Draw a sample of *n* points from the data at random

Fit specified model to that set of *n* points

For each data point outside the sample:

Test the distance from the point to the line against *t*; if the distance from the point to the line is less than *t*, the point is close

end

If there are *d* or more points close to the line then there is a good fit.

Refit the line using all these points.

end

Use the best fit from this collection, using the fitting error as a criterion.

Figure 8. General RANSAC algorithm [from [73]]

We estimate a RANSAC fundamental matrix and then minimize the model estimation errors through a Least Squares Error (LSE) algorithm. The resulting LSE fundamental matrix is used to re-calculate what interest points are inliers and what are not. While the RANSAC generated inliers are subject to the random point selection, meaning the RANSAC-model usually has more error on

the inlier point set in a least-squares sense, the LSE solution on the other hand is a consensus solution over all inliers (instead of just eight) and has smaller error.

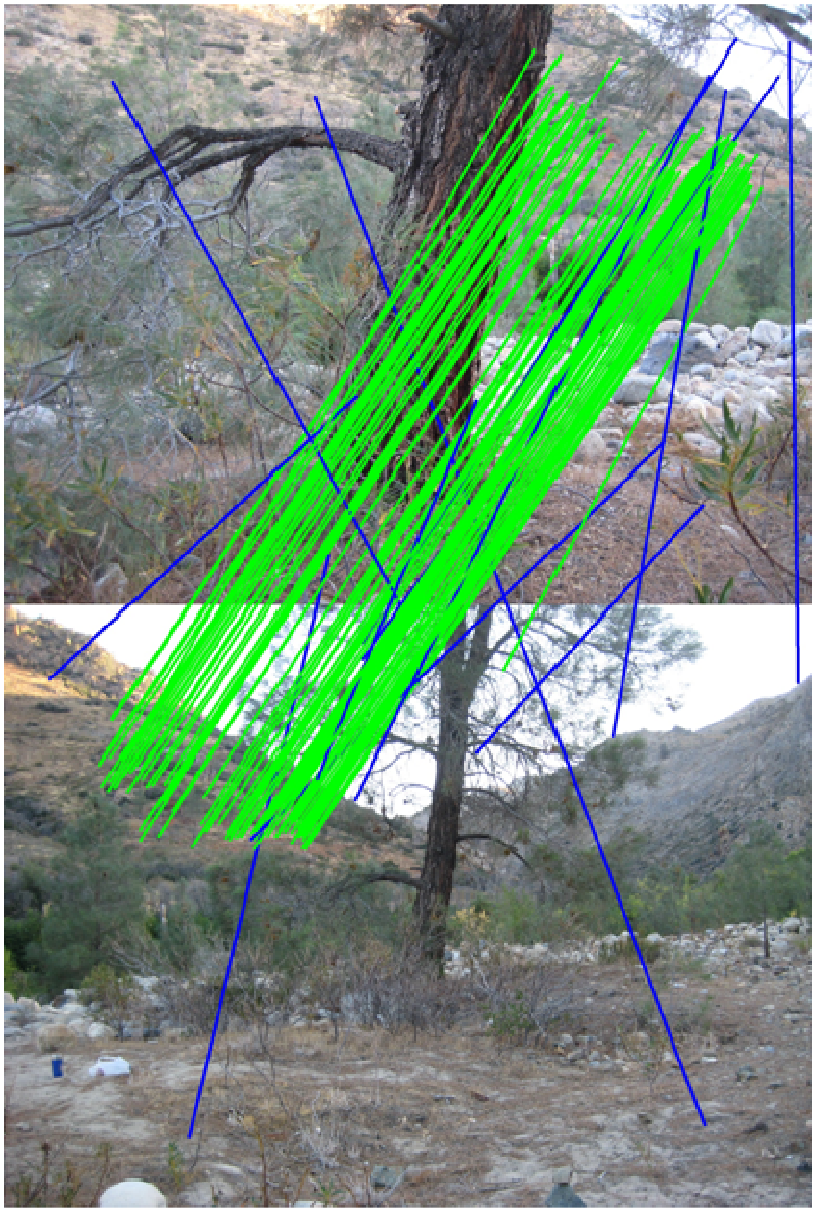


Figure 9. Sample image of matched interest points. Green lines indicate RANSAC inliers, blue lines indicate RANSAC outliers.

We validated this assumption through a manual inspection of the fundamental matrix determined correspondence inliers as shown in Figure 9. If

the visual inspection indicated that the generated ground truth was not accurate, the inliers found in that image were invalidated and the measurements were updated accordingly. These inlier/outlier values are essential to our performance calculations addressed in the next section.

D. PERFORMANCE METRICS

As it is our goal to determine which descriptor offers the best measurable performance within different scene types, we started with precision and recall metrics. Since matching the MSER detected regions to produce a repeatability score (as in [10], [11], [12], and [13]) does not benefit our research, we derived an extractor efficiency score. The measurement scores introduced in this section are generated based upon the partial scene overlap between the scene images and the reference image. We do not calculate the actual overlap region since each extraction technique is evaluated against the same images and each descriptor has equal potential to generate the maximum number of accurate correspondences, the basis of our measurements, within the overlap region.

1. Precision

Precision (or 1-precision) is used in many forms of machine learning research to measure the fidelity of an operation [10], [11], [12]. In this thesis, a high precision score indicates that a high percentage of descriptor-matched points were correctly matched and follow the ground truth fundamental matrix. We calculate precision with the number of true positive matches relative to the total number of descriptor generated correspondence, given by:

$$precision = \frac{\#\,correct\ matches}{\#\,correct\ matches + \#\ false\ matches}$$

1-precision is often plotted verses recall to build a performance curve as presented in [11]. 1-precision is calculated by 1 minus precision.

2. Recall

Equally important as precision is the measurement of an application's ability to correctly operate on all intended elements. A high recall score in this thesis indicates that a descriptor correctly described and matched a high percentage of all possible point correspondences. We calculate recall with the number of true positive matches relative to the total number of possible correspondence given by:

$$recall = \frac{\#correct\ matches}{lessor\ \#of\ detected\ points\ in\ reference\ frame\ or\ image\ frame}$$

Note that this method, by limiting the maximum possible correspondence to the lesser number of detected points in the two images of interest, imposes an upper bound on the measurement of recall.

3. Efficiency

Previous detector and descriptor evaluations presented in [10], [11], [12], [13] utilized a repeatability score introduced by [10]. The repeatability score is primarily a measure of the degree of point detector invariance to transformations such as translation, rotation, illumination, blur and affine. Since we employed only one detector, the repeatability score did not add significant weight to our research. However, in the spirit of what the repeatability score represents for a detector, we developed an efficiency measurement for descriptors. Given that we designed our scene capture pattern in such a fashion as to maximize the potential that each image frame contains the scene imaged in the reference image, each descriptor, based upon the detected regions, has equal opportunity to pick the same regions in the reference image and the subject image.

We calculate efficiency as a measure to which a descriptor was capable of uniquely describing the detected points in both the reference image and the subject image, given by:

$$efficiency = \frac{\#\,correspondence}{lessor\ \#\,of\ detected\ points\ in\ reference\ frame\ or\ image\ frame}$$

E. DESIGN SUMMARY

In this chapter, we have presented our experimental design from the image capture sequence that is utilized to generate our image data set to our specific methods for extracting and matching interest points. Since we captured 99 images at each of seven scenes, our design also includes a novel approach to visualize the results of each experiment run. Additionally, we have explained the performance measures that we will use when presenting our results in the next chapter.

IV. RESULTS AND DISCUSSION

In this chapter, we present the results of our experiments. As defined formally in the previous chapter, we measure performance through precision, recall and efficiency. Precision provides us with a measure of the detail in which an interest point is expressed by a particular descriptor. Recall is a measurement of the relevance of matched points for a particular descriptor. Efficiency measures the ability of a particular descriptor to uniquely encode the detected points. For reference, a perfect descriptor would give a recall of 1.0 and a precision of 1.0 with a high efficiency score relative to the other detectors. Likewise, if a descriptor achieves a high relative efficiency score, but fails to achieve a high precision or recall, then this indicates that while the descriptor was very efficient (i.e. it found many matches between images), few of the matches were correct and therefore the descriptor efficiency would be irrelevant.

Figure 10. Representative data set scene ballroom (“indoor dense”.) The reference image is on the left and another image on the right.

Precision, recall and efficiency scores are calculated for each descriptor for each scene. For example, in our ballroom scene, shown in Figure 10, with the Hess SIFT descriptor, comparing the image captured at an aspect angle of 045, at a distance of 1 meter, and at a camera rotation of 000 against the

reference image yields 10972 interest points in the subject image and 7703 in the reference image. Of these, 1007 were matched based on the Euclidean distance of the integer-based feature vectors with an average distance of 19670. Calculation of the fundamental matrix with RANSAC and LSE produced the following:

$$F = \begin{pmatrix} 2.70032e^{-009} & 4.27533e^{-007} & 0.000172797 \\ 3.392e^{-008} & 1.63211e^{-007} & 0.00309075 \\ -8.57149e^{-006} & 0.00119795 & 1 \end{pmatrix}$$

F was calculated from 964 of the 1007 points. These 964 inliers were within a 1.178 pixel distance of the fundamental matrix generated epipolar lines on the average. The following plot (Figure 11) shows the averaged forward and backward projection position errors for all 1007 points. The distinction between an inlier and an outlier becomes obvious in this plot as the position error is at least one order of magnitude larger for outliers.

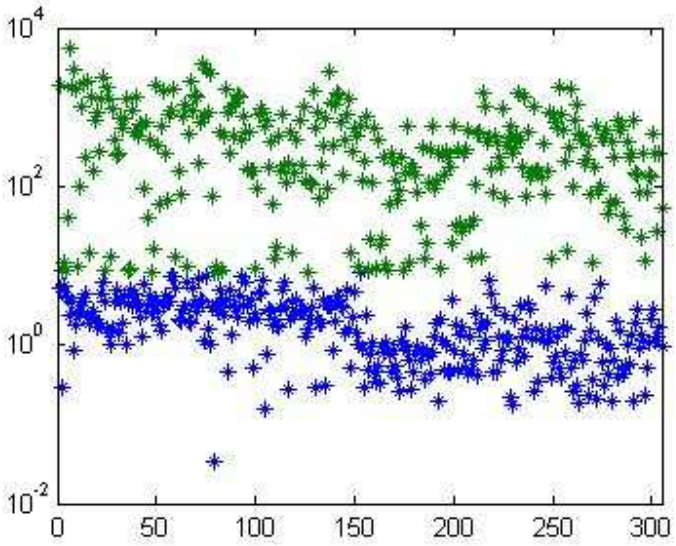


Figure 11. Representative y-axis log scale plot of fundamental matrix re-projection errors. Blue points indicate projections with errors that are eight or fewer pixels, green points indicate errors over eight pixels. Note that our choice of eight pixels is conservative given the inlier and outlier error clustering.

The precision of Hess SIFT for this image is 964/1007 = 0.957299, meaning that 96% of the correspondences were correct matches based on a fundamental matrix. The recall is 964/7703 = 0.125146., meaning that 13% of the possible interest points detected in the reference image were correctly matched by the descriptor. Due to the extreme translation and rotation transformations and the out-of-plane transformations in our scenes, the recall score is expected to be rather small and is used as a relative comparison measure. If we calculated the regions overlap, we could find a theoretical maximum for these values. Efficiency of this detector-descriptor pairing is 1007/7703 = 0.130728, since only 13% of the interest points that were detected in the reference image were found to be matches in the frame image by the descriptor. Again, due to our test sequence, low values in efficiency are normal and expected.

To better visualize the relative performance of each descriptor, we developed a unique, visually appealing method of presenting our results using heat maps. The heat maps are a geometric, top down spatial chart depicting each descriptor's performance at each capture location and image transformation as described in Chapter 3. Figure 12 shows an example of precision, recall, and efficiency, respectively, for all image positions in a spatial heat map.

The descriptor legend digraphs are as follows: HS- Hess SIFT, CF- Complex Filters, GH- GLOH, GM- Gradient Moments, DI- Differential Variants, CC- Cross Correlation, SF- Steerable Filters, PS- PCA SIFT, SC- Shape Context, LS- Lowe SIFT, and SU- SURF. These digraphs are also found in the remaining sections of this chapter. The error bars in Figure 13 show the minimum and maximum values of a particular score in this scene over all camera rotations.

We created cumulative measures for all descriptors, all scenes, all angles, all rotations. Figure 13 is an average over all positions and camera rotations for one scene and for one descriptor.

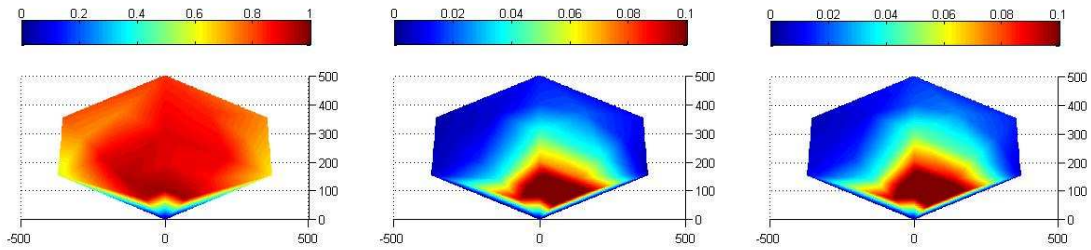


Figure 12. Representative performance heat maps with camera in-plane rotation of 000. The (0,-0.120) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

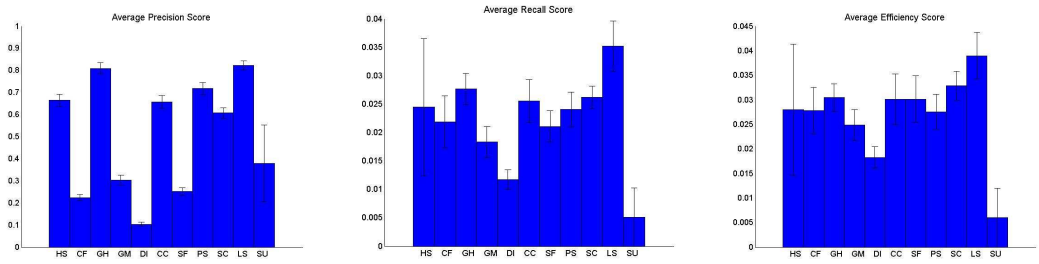


Figure 13. Representative scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.04 and efficiency from 0 to 0.045. The descriptor legend digraphs are as follows: HS- Hess SIFT, CF- Complex Filters, GH- GLOH, GM- Gradient Moments, DI- Differential Variants, CC- Cross Correlation, SF- Steerable Filters, PS- PCA SIFT, SC- Shape Context, LS- Lowe SIFT, and SU- SURF. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

A. INDOOR SCENES

For the purposes of presenting our experiment results, only a representative sample of generated charts is included in this chapter. Appendix A contains a complete compilation of results. The scenes contained in our indoor data set present two indoor environments, one with a dense representation of texture patterns and color palette and one with sparse.

1. Dense: Ballroom

In this scene, we see in Figure 16 that the Lowe SIFT descriptor achieved the highest precision and recall scores, followed closely by GLOH and Hess SIFT. As we can see in the spatial heat maps in Figure 15, Lowe SIFT was also highly invariant to scale and translation transformations. Additionally, the small variance in the precision score indicates that the Lowe SIFT descriptor was also highly invariant to rotation as the error bars are calculated over in-plane image rotations. Notice in Figure 15 how the further distances from the reference image perform worse for all descriptors in terms of recall and efficiency, which is not surprising because of the extreme scale transformations demonstrated by our capture pattern.

The performance scores of the lower dimensional spatial frequency-based descriptors in this scene are significantly lower than the performance scores of the high dimensional, distribution-based SIFT-like descriptors. This seems to indicate that while the spatial frequency methods may perform very well in a short baseline application, their poor performance in a wide-baseline application may preclude usage. The overall low scores in precision and recall for each of our experiments are a direct result of the complex and demanding nature of descriptor matching in our wide baseline test scenes with non-trivial camera transformations.

Of additional note, in a significant portion of the image overlap area with the reference image for all the images captured at a aspect of 22.5 degrees and in the images captured at 135.0 degrees at 4 meters, we observed a bright reflection of the sun on the wall in the vicinity of the center point. We believe that this localized illumination change, contributed to the overall lower performance in the descriptors, except the two SIFT descriptors (each utilizing a different detector) for these images.

Figure 14. Ballroom fountain (dense scene) at distances of 5 meters and aspects of 000 and 045 degrees respectfully, from left to right.

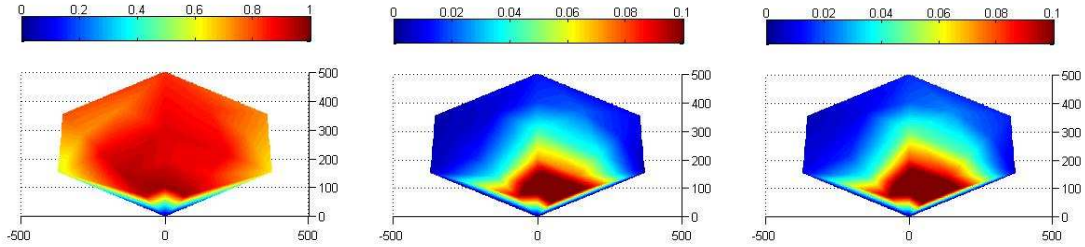


Figure 15. Ballroom scene Lowe SIFT performance heat maps with in-plane camera rotation of 000. The (0,-0.120) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

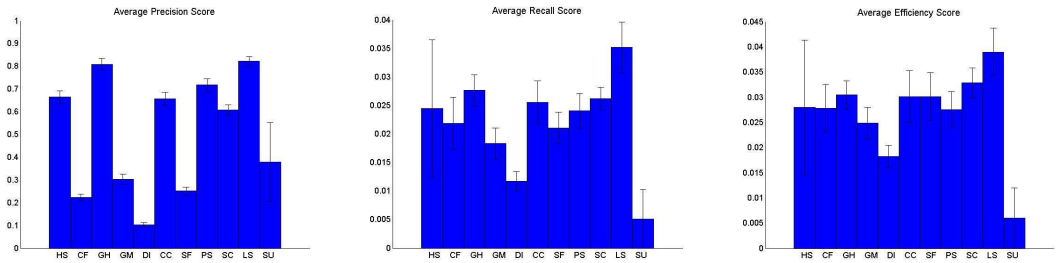


Figure 16. Ballroom scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.04 and efficiency from 0 to 0.045. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

2. Sparse: King Hall

One of our goals in capturing a sparse indoor scene was to demonstrate the relative capability of each descriptor in an environment that is in essence a non-descript white wall with minimal peripheral objects. This scene was captured in a typical foyer of an auditorium on the NPS campus. It was our intuition that few, if any, descriptors within our inventory would provide significant performance scores due to the sparse texture, color and intensity pattern of the foyer.

Not surprisingly, our initial assumptions were validated through this experiment. Here, spatial frequency methods, differential invariants and gradient moments achieved the highest scores in precision and recall followed closely by steerable filters. The corresponding efficiency scores for these descriptors also indicated a higher relative usage of the available detected points. Of special note, only 10 MSER interest points were found in the reference image and an average of 150 points were found in the other images. Upon a closer look, we discovered that all 10 points were found on a small black X that was made with a marker on a piece of tape that we placed on the wall to help center our camera aim point as seen in Figure 20.

The number of interest points detected is a significant deviation from the other scenes (for example, 4432 MSER detected interest points were present in the Ballroom reference image and up to 18,066 were found in other scene images.) Of the 10 interest points in the King Hall reference image, the differential invariant descriptor, for example, found an average of 8 correspondences, the minimum number required for RANSAC to estimate a fundamental matrix. Unsurprisingly, the high dimensional, distribution based descriptors such as SURF, SIFT and GLOH essentially failed to perform any better than the spatial frequency and moment methods in this scene.

Figure 17. King Hall foyer (sparse scene) at distances of 5 meters and aspects of 000 and 045 degrees respectfully, from left to right.

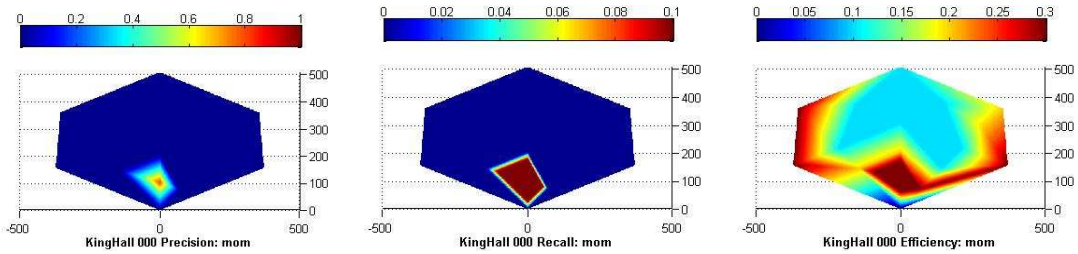


Figure 18. King Hall scene gradient moments performance heat maps with in-plane camera rotation of 000. The (0,-0.120) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

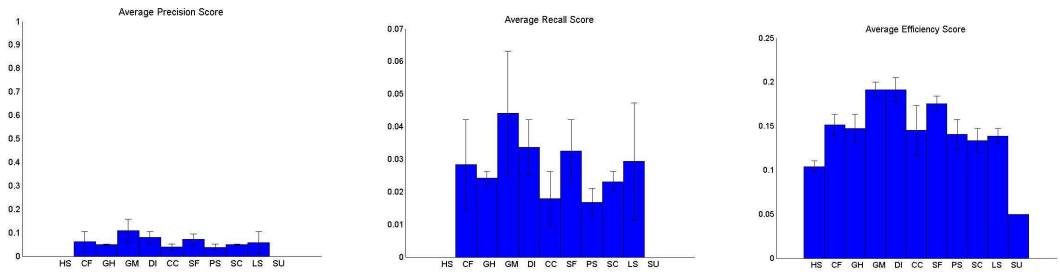


Figure 19. King Hall scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.07 and efficiency from 0 to 0.25. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

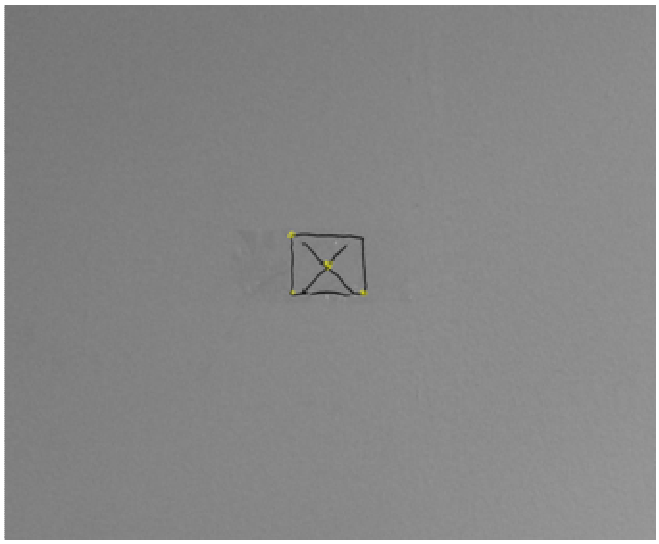


Figure 20. King Hall image captured at an aspect of 090.0 at a distance of 4 meters.

B. OUTDOOR SCENES

In capturing portrayals of outdoor scenes, we desired to demonstrate not only the performance of each descriptor within non-planar, dense scenes of natural texture, but also to demonstrate the environment-relative consistency of our data set collection sequence. As with the indoor scenes, in these experiments we designate the image captured perpendicular and closest to the scene center of each sequence as the sequence reference image.

1. High Desert: Tree

Our first outdoor scene was portrayed in a high-desert classification of environment and was captured in the vicinity of the Fairview national park campground on the Kern River, east of Bakersfield, CA. Note the limited color and texture variations in the scene. These properties are consistent in all our high-desert scenes

The descriptor with the highest average precision and recall scores was the Hess SIFT implementation. Of particular interest in this experiment is the

uniform score distribution of the Hess SIFT precision in the majority of the heat map in Figure 22. This demonstrates a high overall level of invariance to scale and translation transformations in positions with less than 67.5 degrees change in viewpoint.

An important consideration in analyzing all the experiment results is as mentioned in Chapter III, the Hess SIFT (as well as the SURF) descriptor does not use the MSER generated interest points, but instead uses a DoG detector. The high performance scores of the Hess SIFT in this scene may indicate that the use of a detector other than the MSER detector may provide better performance scores for other descriptors in this scene. The wide variances in the Hess SIFT and SURF recall measurements however, indicate that Hess SIFT, or at least the DoG detector it uses, and the SURF descriptor, or the Hessian determinant it uses, are not as invariant to camera rotation transformations as the other descriptors are with the MSER detected interest points.

Figure 21. Tree (high desert scene) at distances of 20 meters and aspects of 000 and 045 degrees respectfully, from left to right.

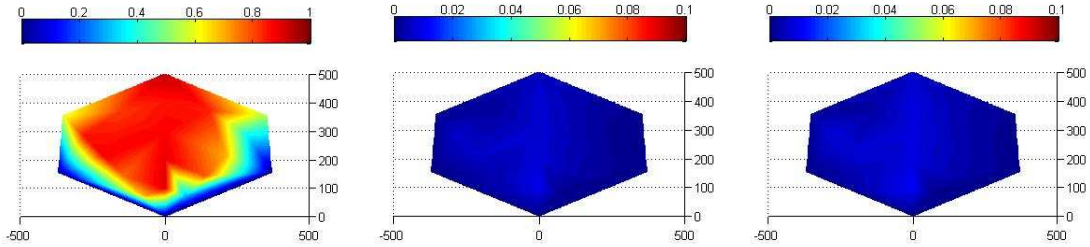


Figure 22. Tree scene Hess SIFT performance heat maps with in-plane camera rotation of 000. The (0,-1) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

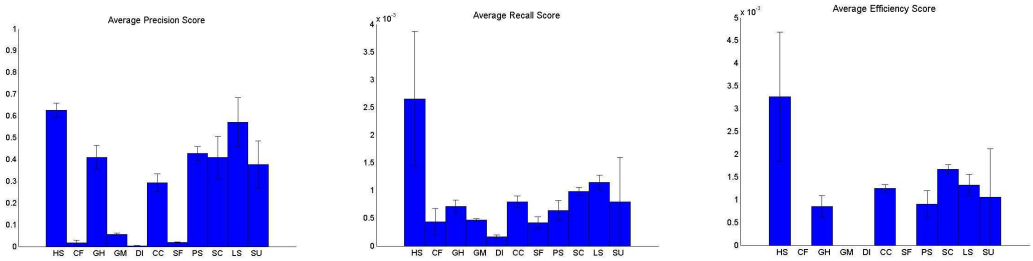


Figure 23. Tree scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.004 and efficiency from 0 to 0.005. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

2. High Desert: Stump

Also captured in the vicinity of Kern River, CA, this scene, as does all our high-desert scenes, presents a natural-occurring color palette and texture pattern with a large metric depth scale and potentially large out-of-plane transformations. Instead of a large tree (which occludes a large portion of the background in each image as with the previous high-desert scene), we centered this scene on the stump of a fallen tree. This approach will induce more non-planar transformations through the change in the distant background.

Once again, in this example of a high desert scene, the Hess SIFT descriptor had the best performance as demonstrated with the highest average precision and recall scores. Hess SIFT also had the highest valid efficiency score. Notice that the highest recall and efficiency scores in Figure 25 are along an axis extending from the center of the scene. This is a result of the large out-of plane transformations inherent in the off-axis capture positions for this scene as well as the unavailability of most of the reference image distant background interest points due to the background regions not being within the field of view of both image capture directions. The other observations made for the previous scene (Figures 24-26 show sample images and results from the High Desert scene) apply to this scene as well.

Figure 24. Stump (high desert scene) at distances of 20 meters and aspects of 000 and 045 degrees respectfully, from left to right.

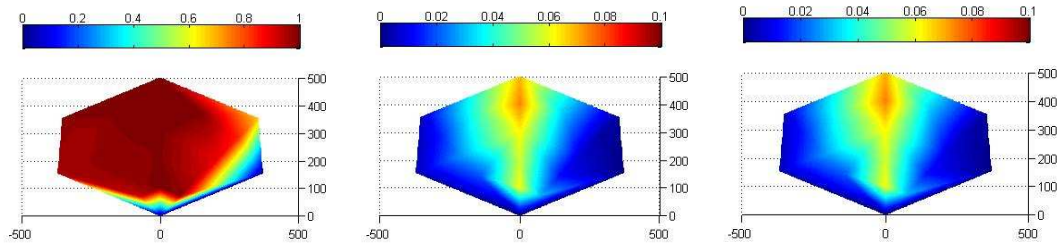


Figure 25. Stump scene Hess SIFT performance heat maps with in-plane camera rotation of 000. The (0,-120) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

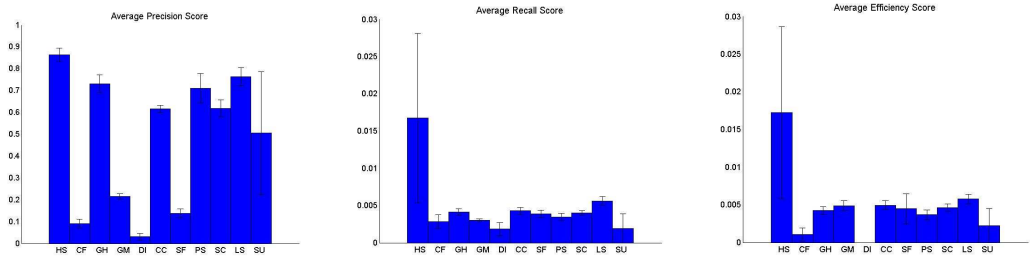


Figure 26. Stump scene performance averaged over in-plane all camera rotations. Precision varies from 0 to 1, recall from 0 to 0.03 and efficiency from 0 to 0.03. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

3. High Desert: Hay Bale

Our final scene captured from a high desert class environment is shown in Figure 27. This scene is centered on a stack of hay bales and presents the same basic scene texture as is found in the previous two scenes; however, this scene also contains a sparse collection of non-natural straight edges in the form of a sign and a fence. As with the previous high-desert scene, this scene will induce more non-planar transformations through the change in the distant background.

Interestingly enough, this scene has generated nearly identical relative performance results as the previous two scenes. This fact demonstrates the consistency and repeatability of the scores generated through our methodology and image capture sequence. Again, notice that the highest recall and efficiency scores in Figure 28 are generated by the images that were captured along an axis extending from the center of the scene.

Of particular note, our analysis of this scene revealed a significant illumination transformation in the images captured at an aspect of 135.0 degrees, 20 meters and an aspect of 157.0 degrees, 16 meters due to the position of the sun. Only the Hess SIFT descriptor was able to correctly match corresponding interest points at these capture positions, for all camera rotations. This suggests a strong invariance of the Hess SIFT descriptor (and/or the DoG detector it employs) to a high level illumination change.

Figure 27. Hay bale (high desert scene) at distances of 20 meters and aspects of 000 and 045 degrees respectfully, from left to right.

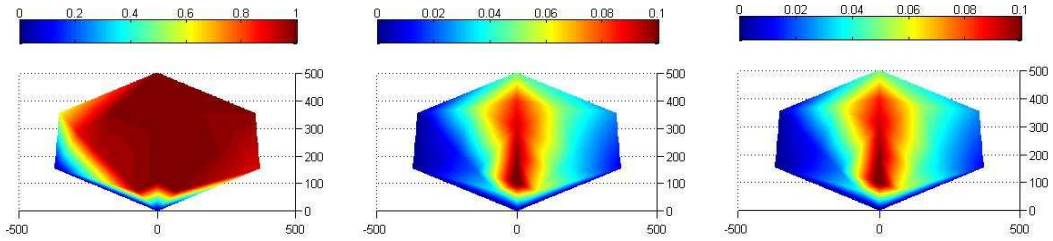


Figure 28. Hay bale scene Hess SIFT performance heat maps with in-plane camera rotation of 000. The (0,-1) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

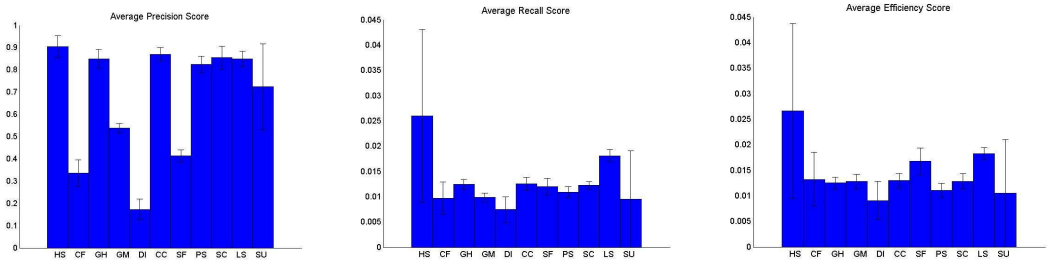


Figure 29. Hay bale scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.045 and efficiency from 0 to 0.045. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

4. Short Building: Halligan Hall

For the short building category in our image data set, we started by imaging a typical 4-story building on the NPS campus. Pictured in Figure 30 is Halligan Hall.

The Hess SIFT descriptor clearly achieved the best relative precision and recall scores, followed closely by SURF. We were surprised at the overall poor performance presented by all the descriptors. Interestingly enough in this scene, the off-axis images that contain both the planar sides of the building demonstrated higher recall scores in Hess SIFT than the complimentary off-axis angles which contain the non-planar distant background (pictured to the right of the building in Figure 30.) This is consistent in SURF, the other relatively high-performing descriptor, and again shows that non-affine transformations pose significant challenges to these types of interest point descriptors.

Figure 30. Halligan Hall (short building scene) at distances of 20 meters and aspects of 000 and 045 degrees respectfully, from left to right.

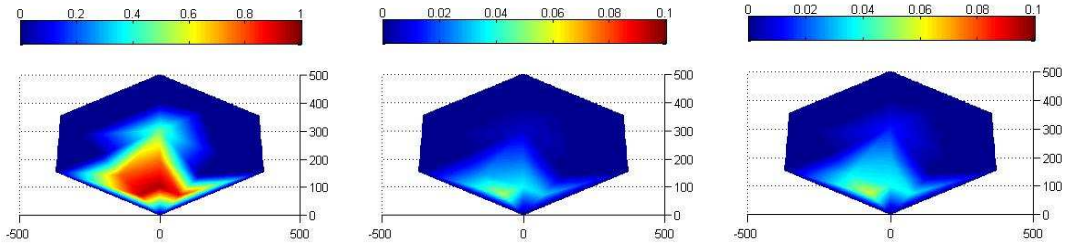


Figure 31. Halligan Hall scene Hess SIFT performance heat maps with in-plane camera rotation of 000. The (0,-1) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

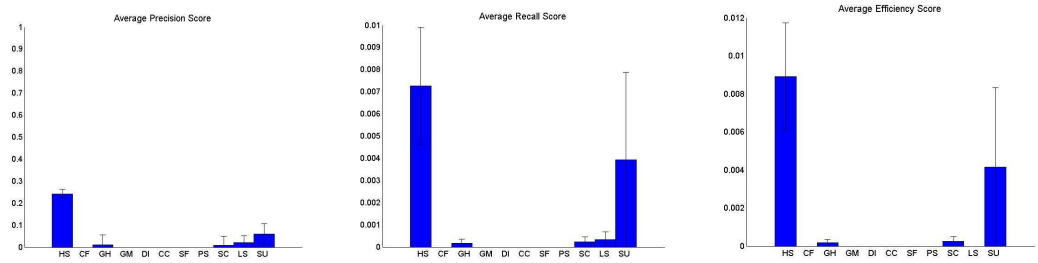


Figure 32. Halligan Hall scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.01 and efficiency from 0 to 0.012. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

5. Short Building: Unmanned Systems Lab

On another side of Halligan Hall on the NPS campus, we captured a scene that presents a different short building aspect. We centered this scene on the front door of the Unmanned Systems Lab.

In this scene, the Lowe SIFT descriptor achieved the highest relative precision/recall paired scores, followed closely by GLOH and PCA SIFT. Similar to the results observed in the previous short building scene, all descriptors had very low relative recall and efficiency scores (note the scales of the performance bar graphs in Figure 35.) This scene includes out-of-plane transformations induced by the outward protrusion of the concrete structure to the left of the door. This caused slightly higher scores in the images where the concrete structure was closer to the center of the images than those where the concrete structure was closer to the outside of the image as shown in Figure 33.

Figure 33. Unmanned Systems Lab (short building scene) at distances of 20 meters and aspects of 000 and 045 degrees respectfully, from left to right.

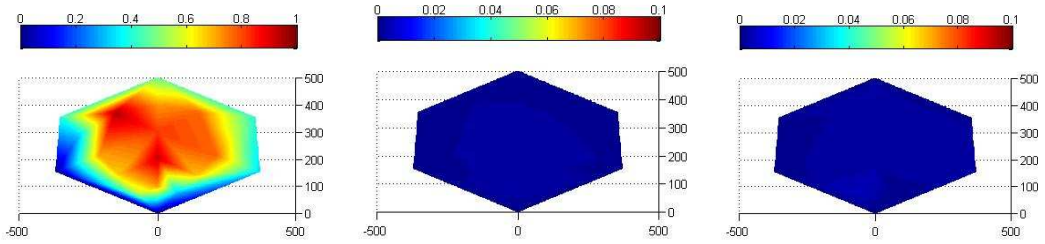


Figure 34. USL scene Lowe SIFT performance heat maps with in-plane camera rotation of 000. The (0,-1) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

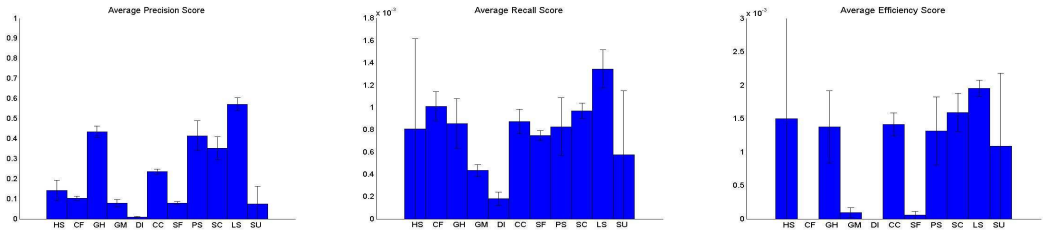


Figure 35. USL scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.0018 and efficiency from 0 to 0.003. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

C. ALTERNATE REFERENCE IMAGE

Through our research analysis, we noticed that no descriptor provided a fair measure of performance in the Halligan Hall and (to a lesser extent) the Unmanned Systems Lab short building scenes. We believe this a direct result of the large scale transformations of the scene combined with the lack of distance background interest points. We replaced the reference image (the image captured perpendicular to and closest to the scene center with the image) with the image captured furthest away from and perpendicular to the scene center. We then re-ran our experiment on both short building scenes.

1. Halligan Hall

We replaced the reference image in the Halligan Hall short-building scene experiment with the image captured at a viewpoint aspect of 090.0 degrees and 20 meters and re-ran the evaluation steps. As we expected, our performance results improved significantly. In Figure 37, we see that the distribution-based methods achieved the highest scores with the Lowe SIFT descriptor slightly outperforming the other high performing methods such as Hess SIFT, GLOH, and Cross Correlation in precision. The Lowe SIFT also performed significantly better then all descriptors in recall and efficiency. Like the experiments that utilized the closest image as the reference image, the positions that captured the scene further away from the reference point performed worse than those captured near the reference image did.

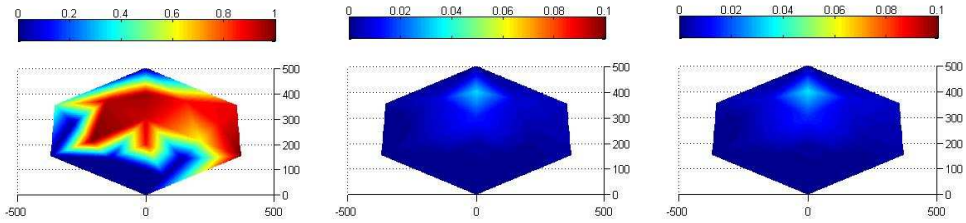


Figure 36. Halligan Hall alternate reference scene Lowe SIFT performance heat maps with in-plane camera rotation of 000. The (0,-1) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

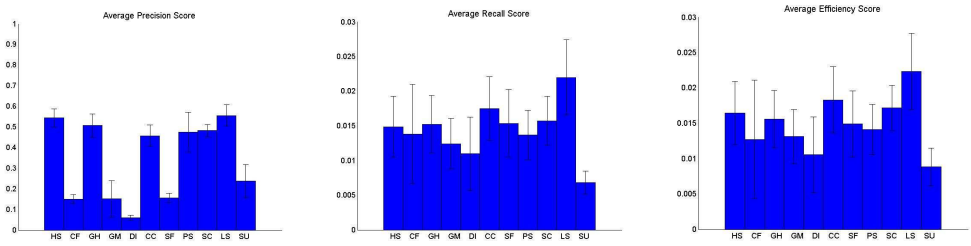


Figure 37. Halligan Hall alternate reference scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.03 and efficiency from 0 to 0.03. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

2. Unmanned Systems Lab

We also re-ran the USL scene experiment with the image captured at a viewpoint aspect of 090.0 degrees and 20 meters as the reference image. With a similar overall performance increase as was seen in the Halligan Hall experiment with the alternate reference image, we have concluded that the choice of reference image does indeed have a significant impact on all the measurements of descriptor performance. As was seen in the previous experiment, Lowe SIFT outperformed all other descriptors.

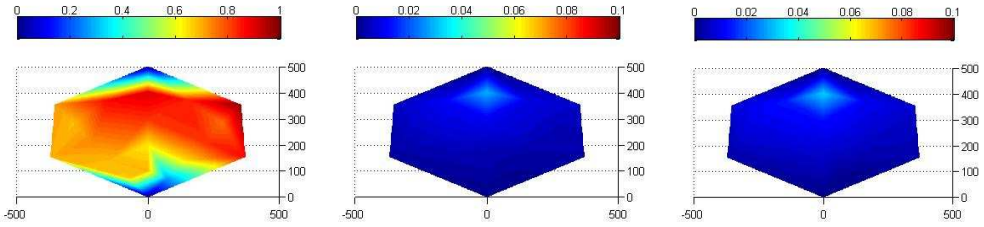


Figure 38. USL alternate reference scene, Lowe SIFT performance heat maps with in-plane camera rotation of 000. The (0,-1) grid location corresponds with the yellow star on the scene image and (0,0) with the focal vertex. Precision (shown in the left graph) improves from 0 to 1, recall (center graph) from 0 to 0.1 and efficiency (right graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)

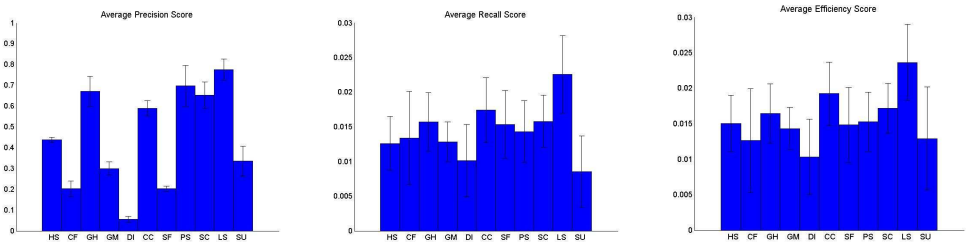


Figure 39. USL alternate reference scene performance averaged over all in-plane camera rotations. Precision varies from 0 to 1, recall from 0 to 0.03 and efficiency from 0 to 0.03. Descriptor digraphs are the same as defined in Figure 13. Error bars indicate the minimum and maximum of the measurement among all camera rotations.

D. OVERALL RESULT DISCUSSION

Given the complex nature of the appearance changes achieved through the camera transformations demonstrated by our capture pattern, we were highly impressed with the individual performance of every descriptor tested within our inventory. We did not expect to find that every descriptor produces primarily accurate Fundamental matrices for almost all image sequences. Although each descriptor did surprisingly well, there were clear and distinctive top performers. All the SIFT-based descriptors obtain the best results in each scene except the King Hall (sparse scene) where no descriptor had sufficient performance to generate a reliable fundamental matrix. This might be a result of the application

of the MSER detected interest points for each descriptor. It is quite possible that the non-distribution-based descriptors require interest points found by other attributes than those found by MSER to operate properly. While this aspect is beyond the scope of this thesis, we are intrigued by the possibility that the detector/descriptor pairing choice might be the most beneficial study given our research interests.

The performance scores of the MSER-based Lowe SIFT and other SIFT-derived extractors and the DoG-based Hess SIFT extractor were impressive. Specifically, the GLOH and Hess SIFT descriptors typically only correlated interest points (near zero percentage of outliers to inliers) that would later find accurate fundamental matrices. This occurred even in scenes where other descriptors could not find an accurate fundamental matrix. Our analysis of the high-desert Hay Bale scene seems to indicate a strong invariance of the Hess SIFT descriptor (and/or the DoG detector it employs) to a high-level illumination change. We also found that in the Ballroom indoor scene, the Hess SIFT and SURF recall measurements incurred a wide variation throughout the camera rotations. This indicates that Hess SIFT, (again and/or the DoG detector it employs,) and the SURF descriptor, (and/or the Hessian determinant it employs,) are not as invariant to camera rotation transformations as the other descriptors are with interest points detected by MSER.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND FUTURE WORK

A. SUMMARY

In this thesis, we have presented a novel evaluation method for vision-based interest point extraction that bases relative performance merit upon the level of demonstrated invariance within a scene type. We have also presented a comparison of our performance measures with the conventional approaches introduced by [11].

In general, we found the use of high-resolution images very time-intensive and processing resource demanding. Our experimental runs for all detectors and all scenes took on average 24 hours to complete on our dual Intel Xeon quad core processor test platform with 16 GB of RAM. We also found that compiling the results in the form of heat maps, plots, and stacked images with line correspondences was very taxing on available hard drive space. Our total compilation of results occupied over 100 GB of hard drive space.

The following paragraphs address our research questions individually.

1. Extractor Selection based on Scene Classifications

It was our goal to determine an appropriate interest point extractor to apply to an environment-based scene classification. We specifically focused on interest point description, the second step of extraction, to be a key determination of classification-based performance. Our research did not utilize a sufficient number of representative scenes of each class to draw decisive conclusions. It has shown that while each descriptor produced a clear level of performance in each scene type tested, the high-dimensional scale and rotation invariant descriptors (SIFT, GLOH, and sometimes SURF) consistently outperformed the other methods. The fact that the relative performance scores of each descriptor were nearly identical in each of the three scene representatives of a high desert

scene seems to indicate that our test framework can consistently and quantifiable determine a best-suited extractor. We found that the MSER-based, Lowe SIFT extractor produced the highest scores in the dense indoor scene represented by the Ballroom images. We also found that the Lowe SIFT extractor performed the best in the two short building scenes when we changed our reference image. Additionally, the observation that the same SIFT-derived extractor (Hess SIFT) consistently produced the highest scores within the three high desert scenes, indicates that perhaps the key to conclusive results in this research is to test only SIFT-derived methods and couple them to different detectors to measure the overall extractor performance in a given scene type.

2. Multiple Extractor Employment within a Single Image.

Ultimately, in an autonomous system platform, we would like to be able to sub-segment an image into semantically relevant regions and then employ the most suitable extractors in each. This research only provides the initial experimental results of performance-based pairing of an extraction technique to a scene classification. Given that, we believe that while processing resource limitations such as space and latency may preclude autonomous system use of this method of multiple extractor employment, our results indicate that given accurate and consistent sub-scene classification, it should be possible to optimally employ the best extractor.

B. FUTURE WORK

In the process of conducting this work, we came across many additional interesting avenues of inquiry than we were unable to conduct in the allotted time. For example, we were unable to exhaustively test as many scene types as we desired. We captured scenes representative of indoor sparse, indoor dense, outdoor high desert and outdoor short building. Future work should include image classifications such as office, living room, bedroom, kitchen, store,

industrial, tall building, inside city, highway, coast, open country, mountain, forest and suburb scene classifications buildings, city/street, agriculture/countryside, desert, and sea.

In addition, inclusion of more examples of each scene type should be considered. While the three high-desert scenes provide a measure of scene consistency for this research, for conclusive analysis, more scenes would be required for each classification type.

In this research, we decided to assign the closest, perpendicular (aspect of 090.0) image to the scene center as our reference image for both indoor and outdoor classifications. As we discovered in our short building scenes, this may not be the best method to analyze descriptor performance. In future experiments, specifically in an outdoor environment with a much larger scale environment with natural occurring texture, consideration should be given to choosing an alternate image to designate as the reference image.

Each detector and descriptor contains a set of operating parameters that allow fine-tuning of the algorithm to better fit the employment domain. We tested each of the extraction pairs with the author recommended settings. It would be an interesting effort to individually vary each parameter of each detector and each parameter of each descriptor to determine which paring of extractors perform the best with which set of paired parameters. However, this complete study would be extremely, if not prohibitively, computationally intensive.

Our efforts primarily employed only one detector for each descriptor with the exception of Hess SIFT and SURF. It is our opinion that the spatial frequency methods and the gradient methods might perform better when coupled to different detectors.

Our method of calculating camera motion ground truth performed well in our planar and non-planar scenes. However, the employment of additional ground truth finding methods such as the Trifocal Tensor method described in Chapter II, or employing an average function of all the scene fundamental

matrices generated for a camera capture sequence might be beneficial to help evaluate the results of images with large out-of-plane transformations. Another option would be to calculate the true fundamental matrix given the actual camera movement and the calibration matrix.

We presented a novel approach to the comparison of interest point extractors. It would be interesting to compare our method of producing measurement scores to the scores generated with the conventional method found in [11]. The comparison would provide an opportunity to contrast the benefits of the two approaches.

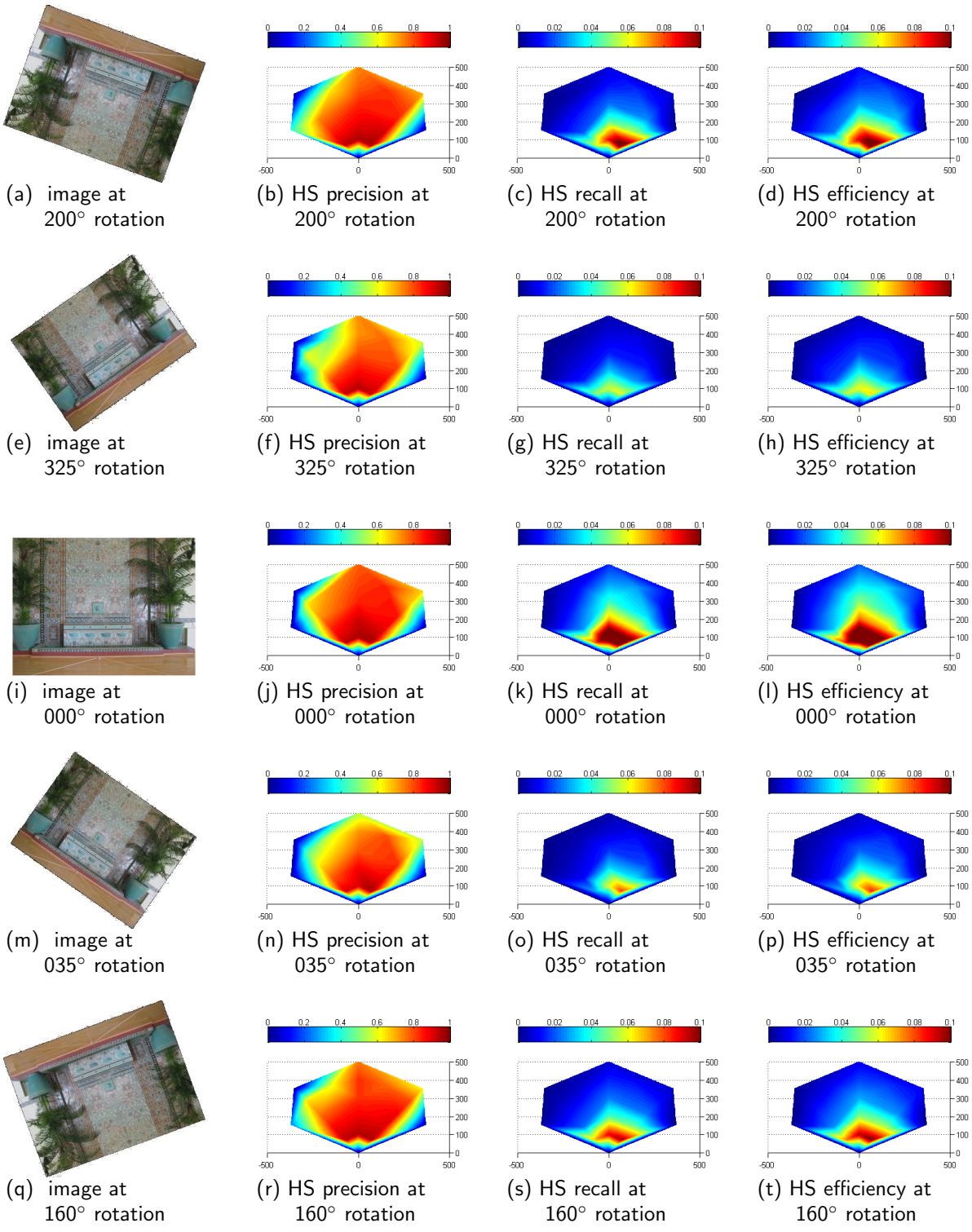
As we stated in Chapter I, our research interests are in autonomous systems. With that said, our research did not seek to explore the clearly important issues of processing and memory limitations or time latency considerations of each descriptor. This information would be required prior to properly determining which interest point extraction solution to deploy in a vision-enabled autonomous system. Most autonomous systems also only have lower resolution images available.

C. CONCLUSION

The results of this research provide immediate benefit to the current and future projects of the NPS Vision lab by facilitating the utilization of the best-suited extraction techniques. Furthermore, the increased reliance of the United States armed forces on the standoff war-fighting capabilities of unmanned and autonomous vehicles (UXV) in, on, and above the sea, necessitates better overall navigation capabilities of these platforms. Our research contributes an important cornerstone towards the validation of precision, vision-based navigation, thereby increasing UXV performance and strengthening the security of the United States and her allies worldwide.

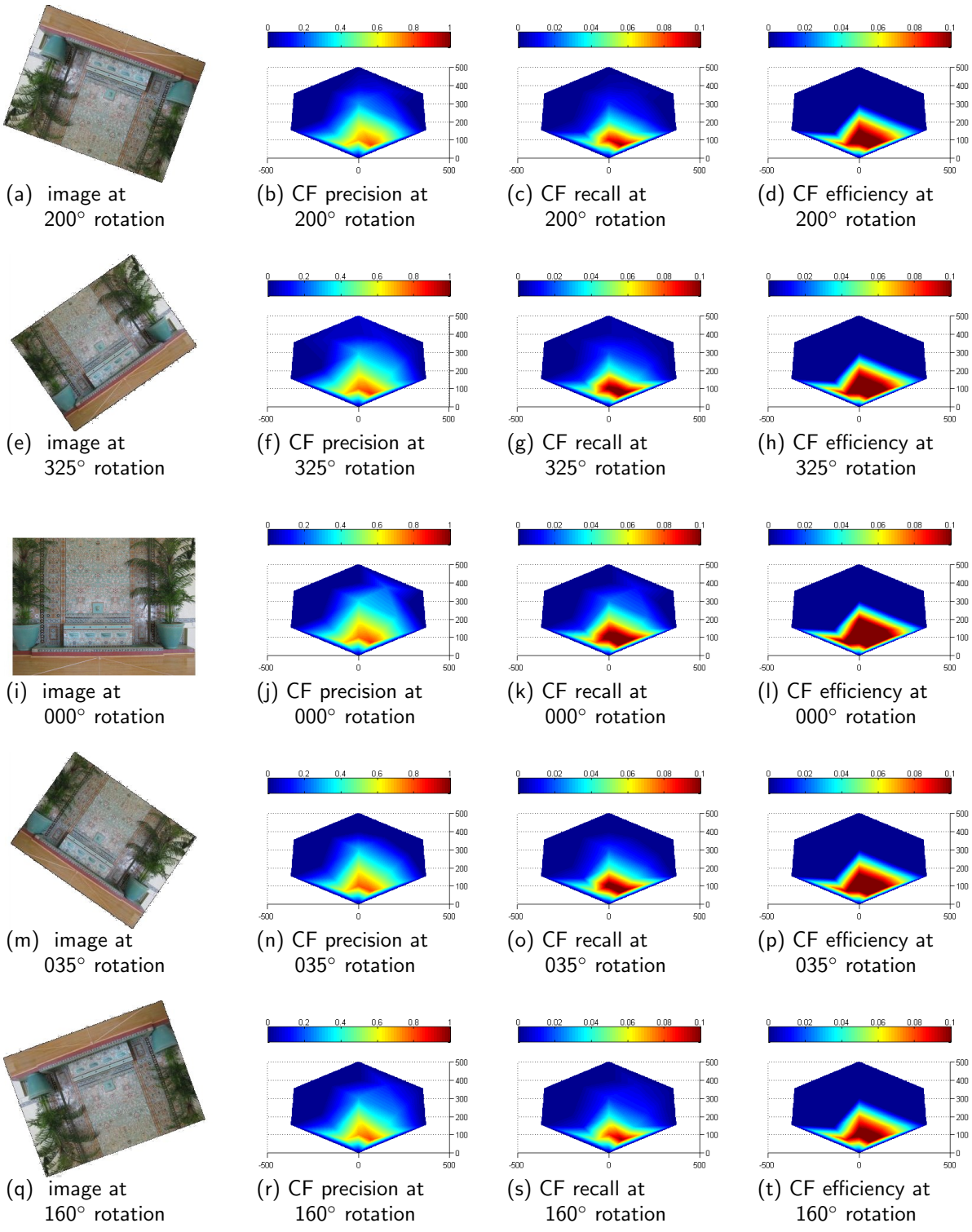
APPENDIX A

This appendix contains the spatial heat maps for each scene, for each position, for each camera rotation produced as a result of our research. Each scene section is preceded by an example image of the scene. The descriptors are identified by a two letter digraphs and are defined as follows: HS- Hess SIFT, CF- Complex Filters, GH- GLOH, GM- Gradient Moments, DI- Differential Variants, CC- Cross Correlation, SF- Steerable Filters, PS- PCA SIFT, SC- Shape Context, LS- Lowe SIFT, and SU- SURF. Precision (shown in the left heat map graph) improves from 0 to 1, recall (center heat map graph) from 0 to 0.1 and efficiency (right heat map graph) from 0 to 0.1 as indicated in each map as blue (cold, bad) to red (hot, good.)



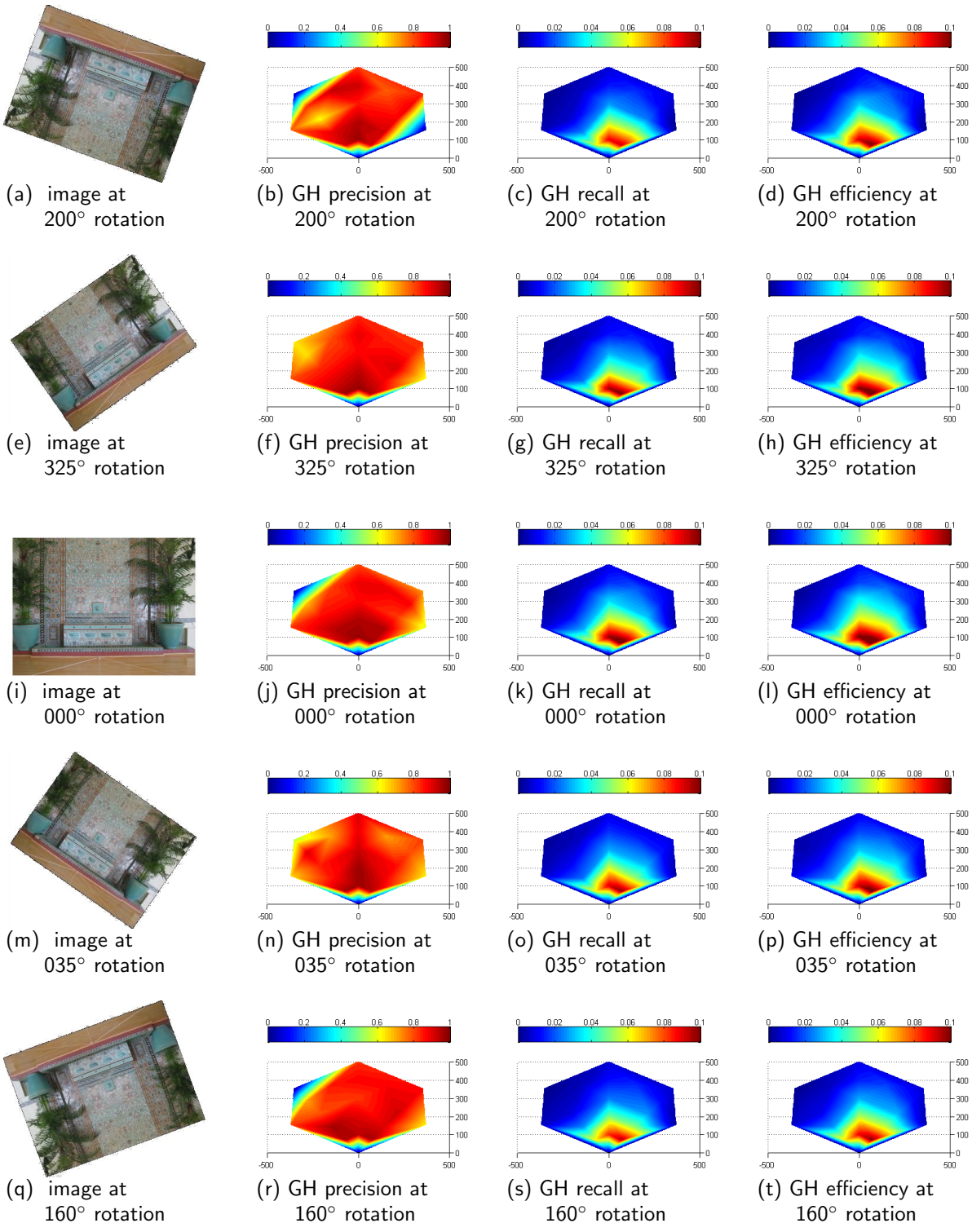
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 40. Heat maps for descriptor HS in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



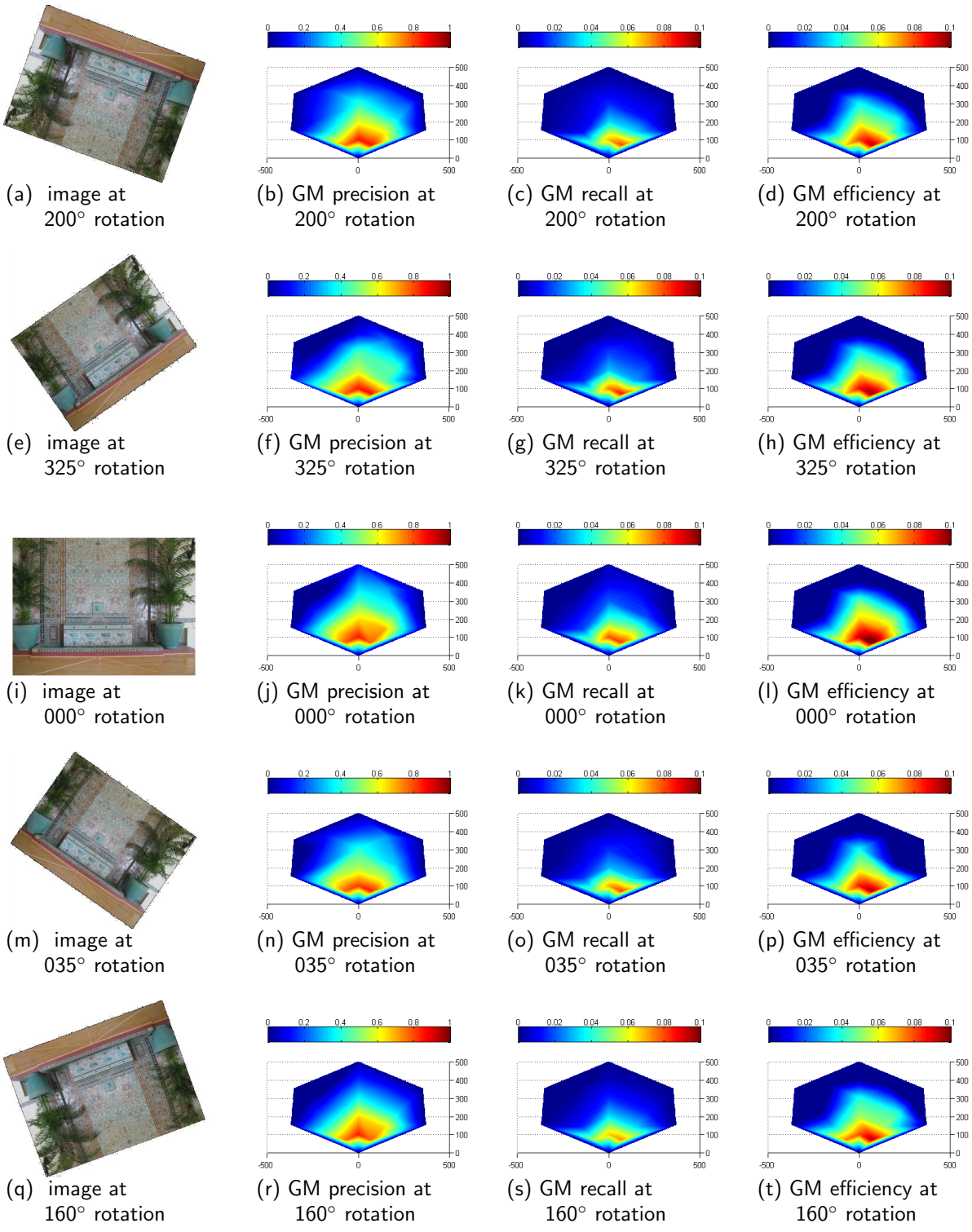
(a) image at 200° rotation
(b) CF precision at 200° rotation
(c) CF recall at 200° rotation
(d) CF efficiency at 200° rotation
(e) image at 325° rotation
(f) CF precision at 325° rotation
(g) CF recall at 325° rotation
(h) CF efficiency at 325° rotation
(i) image at 000° rotation
(j) CF precision at 000° rotation
(k) CF recall at 000° rotation
(l) CF efficiency at 000° rotation
(m) image at 035° rotation
(n) CF precision at 035° rotation
(o) CF recall at 035° rotation
(p) CF efficiency at 035° rotation
(q) image at 160° rotation
(r) CF precision at 160° rotation
(s) CF recall at 160° rotation
(t) CF efficiency at 160° rotation

Figure 41. Heat maps for descriptor CF in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



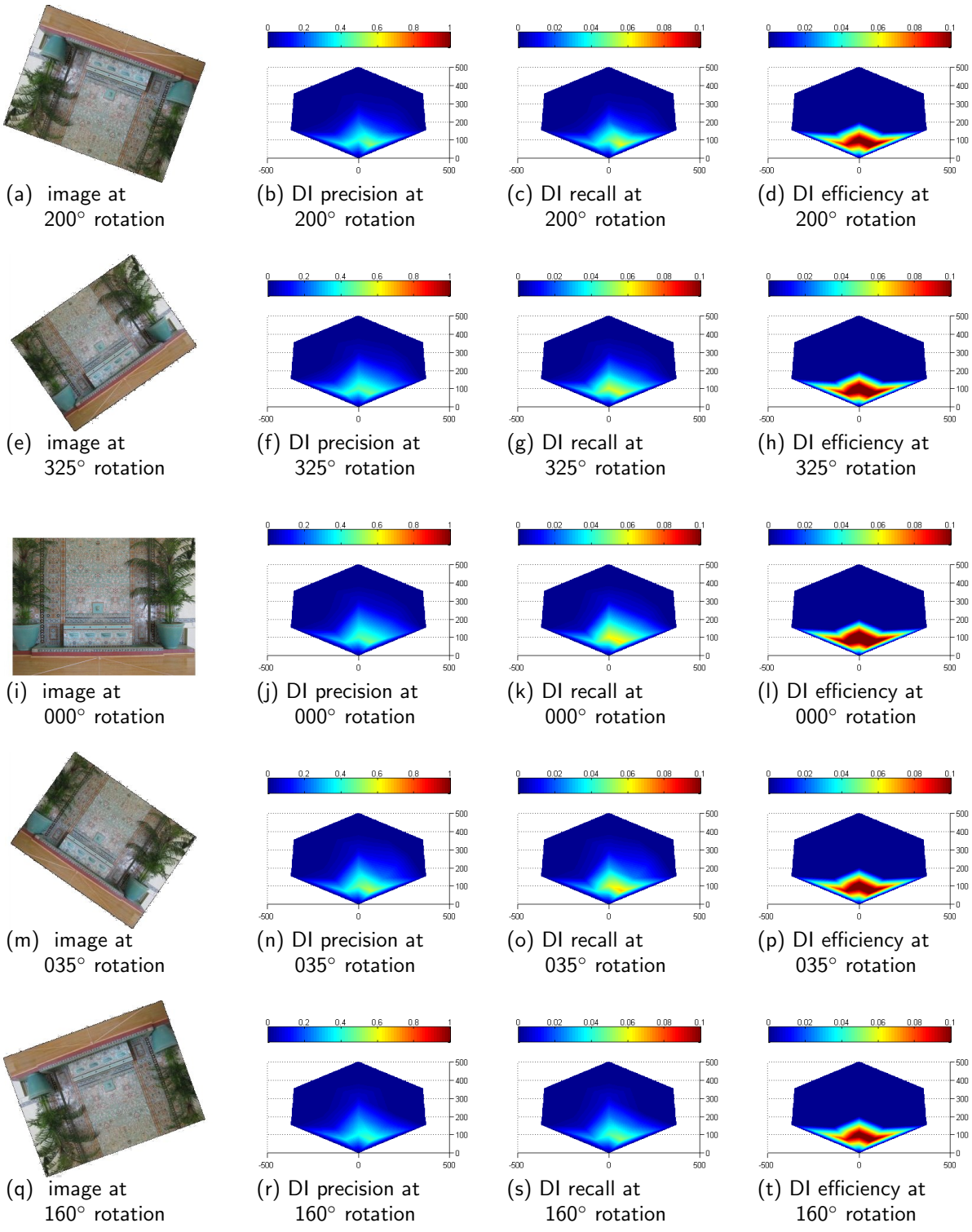
(a) image at 200° rotation
(b) GH precision at 200° rotation
(c) GH recall at 200° rotation
(d) GH efficiency at 200° rotation
(e) image at 325° rotation
(f) GH precision at 325° rotation
(g) GH recall at 325° rotation
(h) GH efficiency at 325° rotation
(i) image at 000° rotation
(j) GH precision at 000° rotation
(k) GH recall at 000° rotation
(l) GH efficiency at 000° rotation
(m) image at 035° rotation
(n) GH precision at 035° rotation
(o) GH recall at 035° rotation
(p) GH efficiency at 035° rotation
(q) image at 160° rotation
(r) GH precision at 160° rotation
(s) GH recall at 160° rotation
(t) GH efficiency at 160° rotation

Figure 42. Heat maps for descriptor GH in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



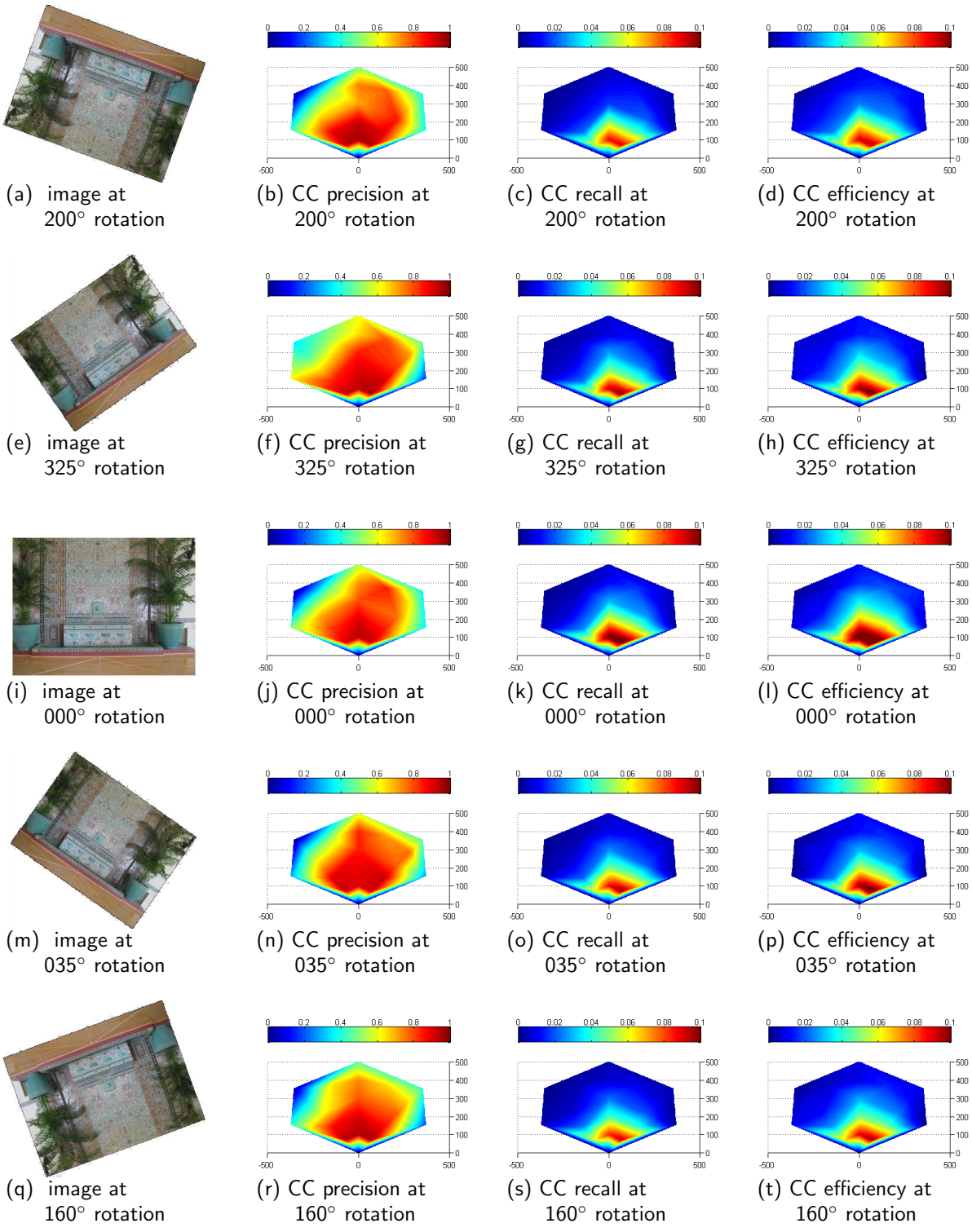
(a) image at 200° rotation
(b) GM precision at 200° rotation
(c) GM recall at 200° rotation
(d) GM efficiency at 200° rotation
(e) image at 325° rotation
(f) GM precision at 325° rotation
(g) GM recall at 325° rotation
(h) GM efficiency at 325° rotation
(i) image at 000° rotation
(j) GM precision at 000° rotation
(k) GM recall at 000° rotation
(l) GM efficiency at 000° rotation
(m) image at 035° rotation
(n) GM precision at 035° rotation
(o) GM recall at 035° rotation
(p) GM efficiency at 035° rotation
(q) image at 160° rotation
(r) GM precision at 160° rotation
(s) GM recall at 160° rotation
(t) GM efficiency at 160° rotation

Figure 43. Heat maps for descriptor GM in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



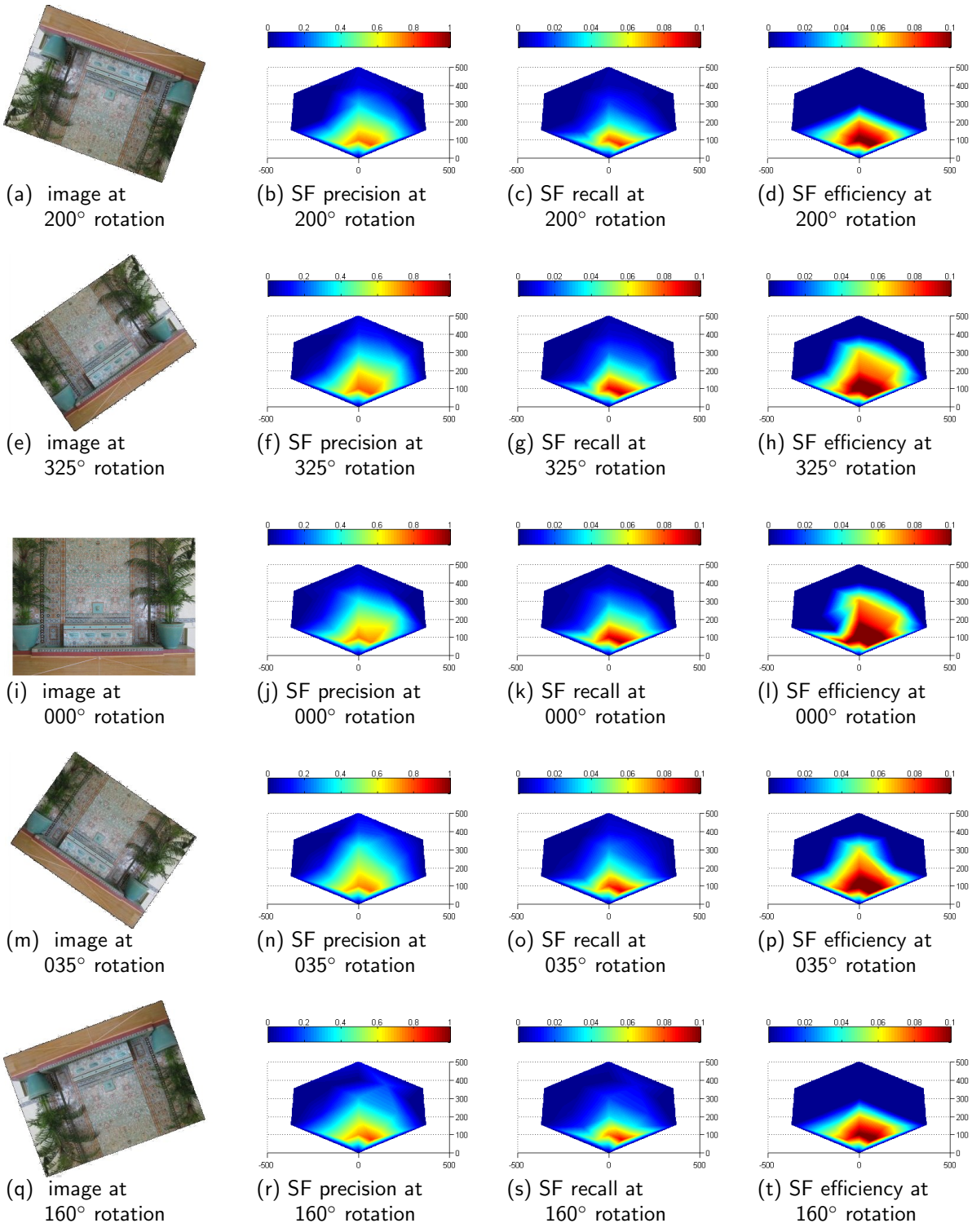
(a) image at 200° rotation
(b) DI precision at 200° rotation
(c) DI recall at 200° rotation
(d) DI efficiency at 200° rotation
(e) image at 325° rotation
(f) DI precision at 325° rotation
(g) DI recall at 325° rotation
(h) DI efficiency at 325° rotation
(i) image at 000° rotation
(j) DI precision at 000° rotation
(k) DI recall at 000° rotation
(l) DI efficiency at 000° rotation
(m) image at 035° rotation
(n) DI precision at 035° rotation
(o) DI recall at 035° rotation
(p) DI efficiency at 035° rotation
(q) image at 160° rotation
(r) DI precision at 160° rotation
(s) DI recall at 160° rotation
(t) DI efficiency at 160° rotation

Figure 44. Heat maps for descriptor DI in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



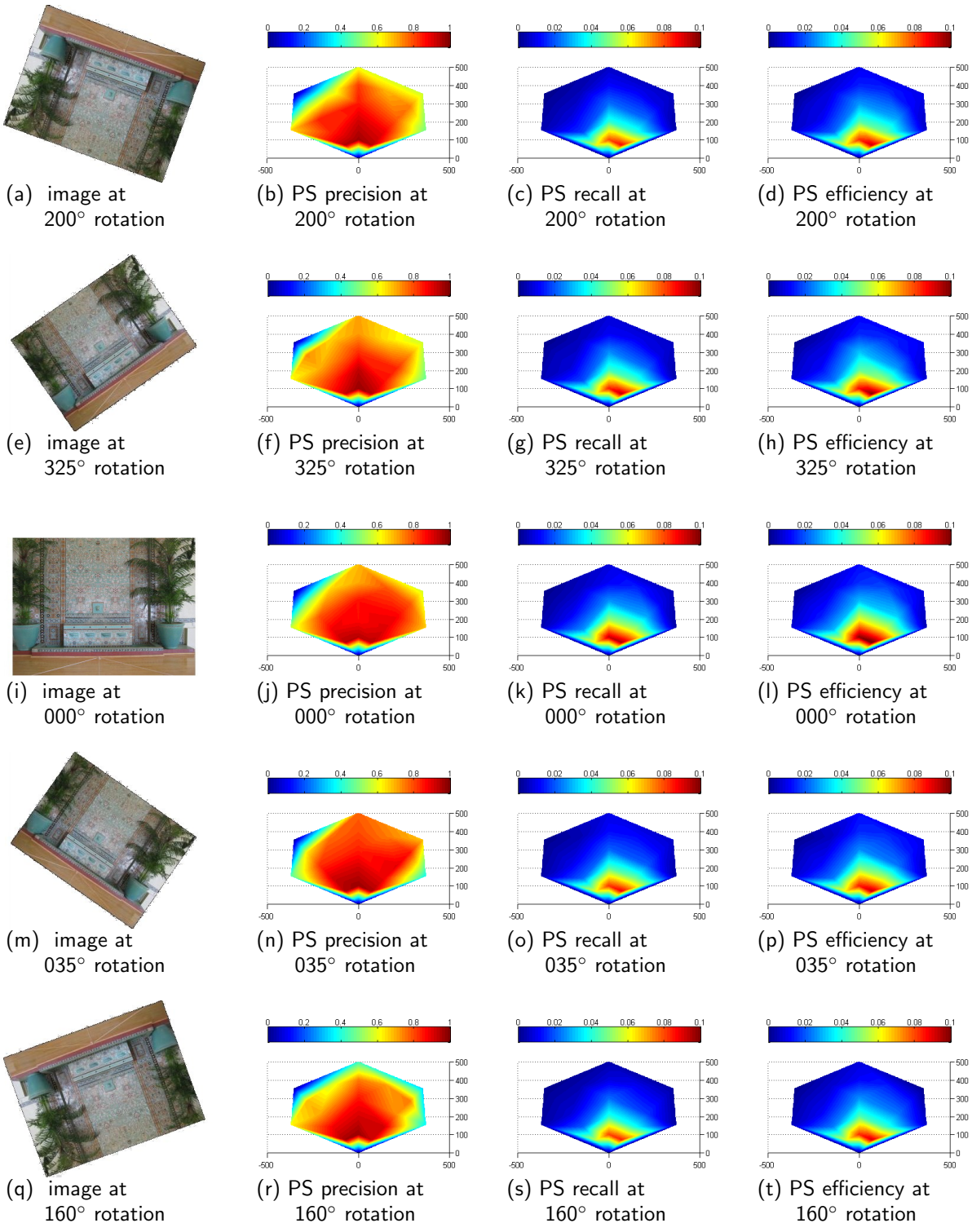
(a) image at 200° rotation
(b) CC precision at 200° rotation
(c) CC recall at 200° rotation
(d) CC efficiency at 200° rotation
(e) image at 325° rotation
(f) CC precision at 325° rotation
(g) CC recall at 325° rotation
(h) CC efficiency at 325° rotation
(i) image at 000° rotation
(j) CC precision at 000° rotation
(k) CC recall at 000° rotation
(l) CC efficiency at 000° rotation
(m) image at 035° rotation
(n) CC precision at 035° rotation
(o) CC recall at 035° rotation
(p) CC efficiency at 035° rotation
(q) image at 160° rotation
(r) CC precision at 160° rotation
(s) CC recall at 160° rotation
(t) CC efficiency at 160° rotation

Figure 45. Heat maps for descriptor CC in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



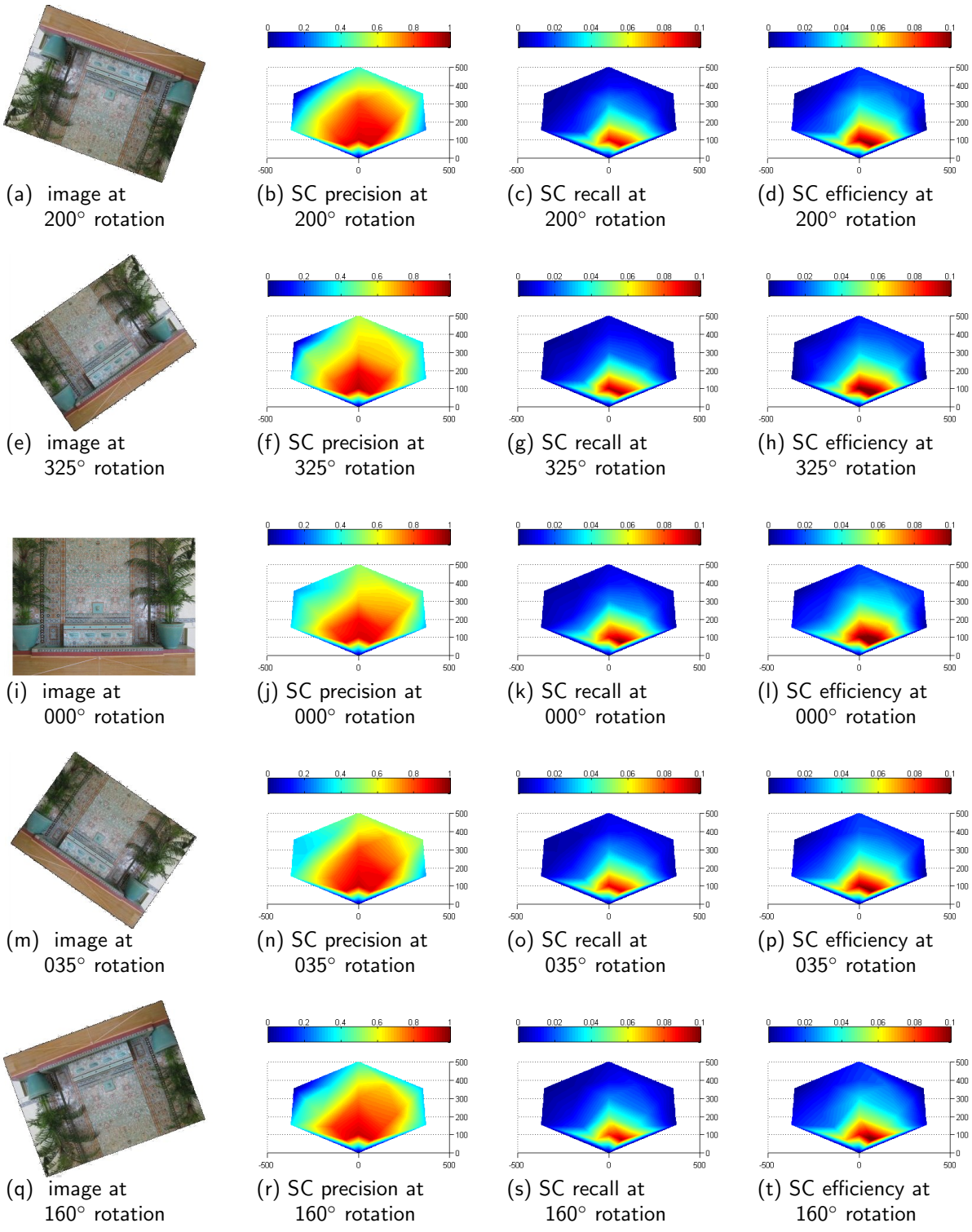
(a) image at 200° rotation
(b) SF precision at 200° rotation
(c) SF recall at 200° rotation
(d) SF efficiency at 200° rotation
(e) image at 325° rotation
(f) SF precision at 325° rotation
(g) SF recall at 325° rotation
(h) SF efficiency at 325° rotation
(i) image at 000° rotation
(j) SF precision at 000° rotation
(k) SF recall at 000° rotation
(l) SF efficiency at 000° rotation
(m) image at 035° rotation
(n) SF precision at 035° rotation
(o) SF recall at 035° rotation
(p) SF efficiency at 035° rotation
(q) image at 160° rotation
(r) SF precision at 160° rotation
(s) SF recall at 160° rotation
(t) SF efficiency at 160° rotation

Figure 46. Heat maps for descriptor SF in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



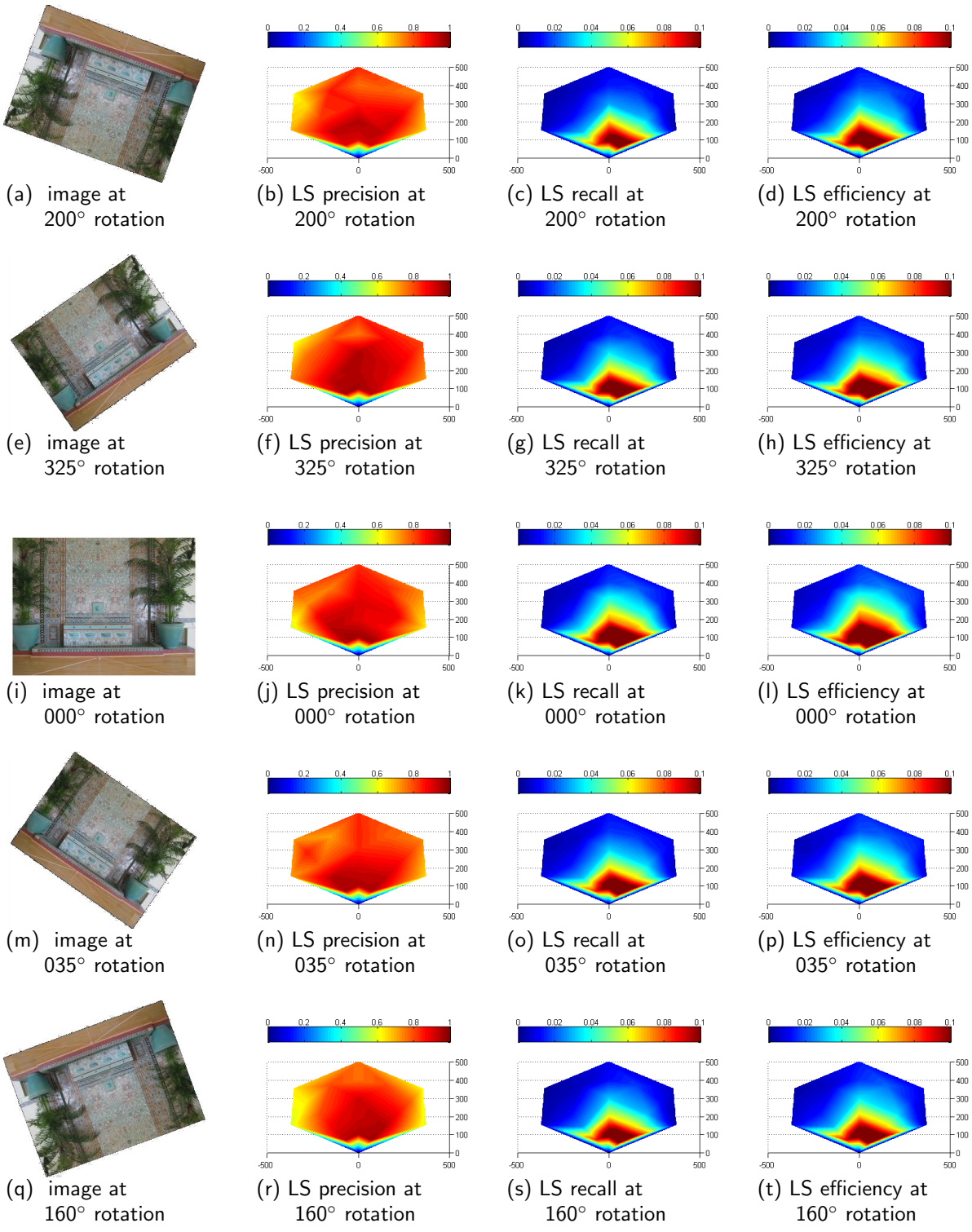
(a) image at 200° rotation
(b) PS precision at 200° rotation
(c) PS recall at 200° rotation
(d) PS efficiency at 200° rotation
(e) image at 325° rotation
(f) PS precision at 325° rotation
(g) PS recall at 325° rotation
(h) PS efficiency at 325° rotation
(i) image at 000° rotation
(j) PS precision at 000° rotation
(k) PS recall at 000° rotation
(l) PS efficiency at 000° rotation
(m) image at 035° rotation
(n) PS precision at 035° rotation
(o) PS recall at 035° rotation
(p) PS efficiency at 035° rotation
(q) image at 160° rotation
(r) PS precision at 160° rotation
(s) PS recall at 160° rotation
(t) PS efficiency at 160° rotation

Figure 47. Heat maps for descriptor PS in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



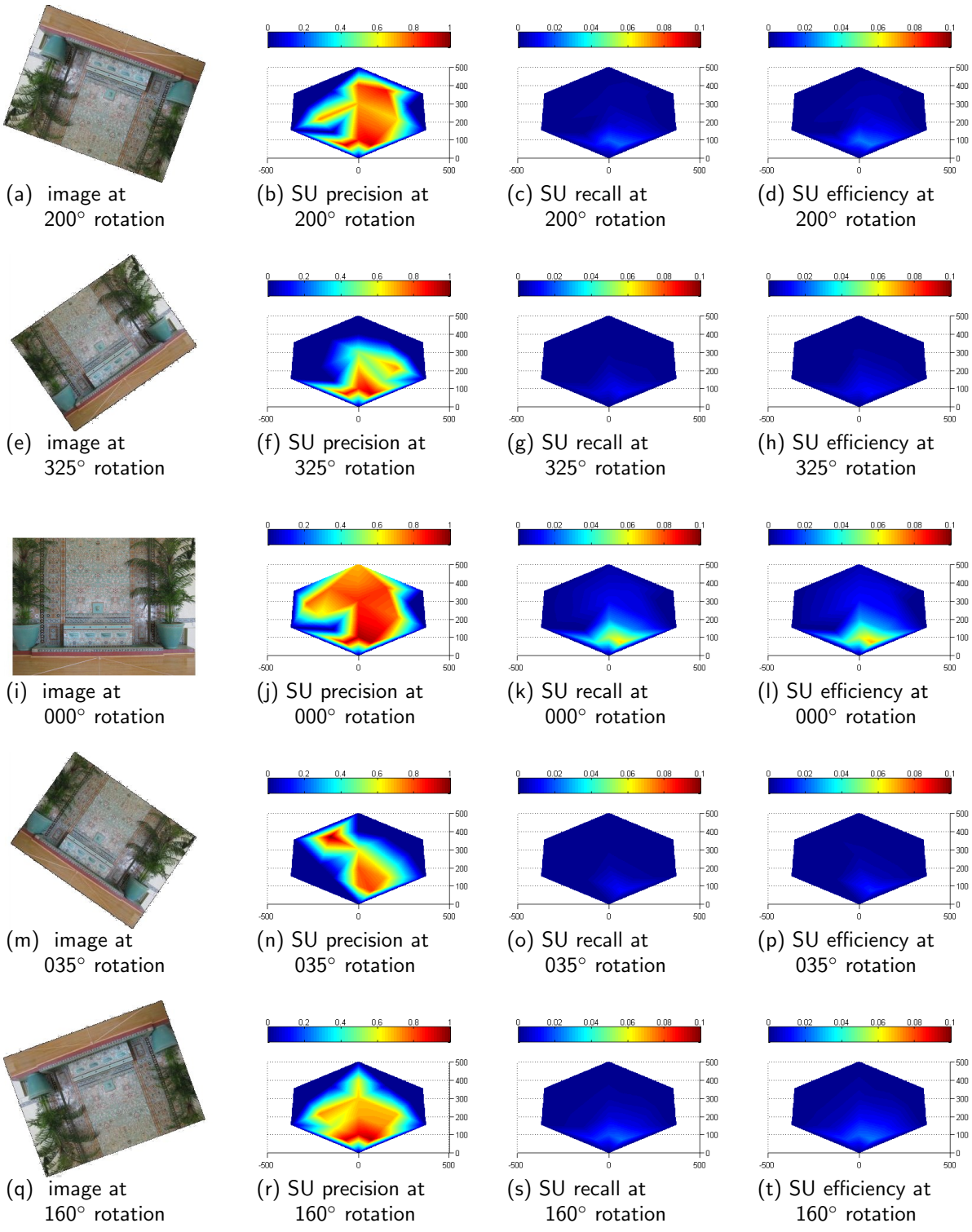
(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

Figure 48. Heat maps for descriptor SC in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



(a) image at 200° rotation
(b) LS precision at 200° rotation
(c) LS recall at 200° rotation
(d) LS efficiency at 200° rotation
(e) image at 325° rotation
(f) LS precision at 325° rotation
(g) LS recall at 325° rotation
(h) LS efficiency at 325° rotation
(i) image at 000° rotation
(j) LS precision at 000° rotation
(k) LS recall at 000° rotation
(l) LS efficiency at 000° rotation
(m) image at 035° rotation
(n) LS precision at 035° rotation
(o) LS recall at 035° rotation
(p) LS efficiency at 035° rotation
(q) image at 160° rotation
(r) LS precision at 160° rotation
(s) LS recall at 160° rotation
(t) LS efficiency at 160° rotation

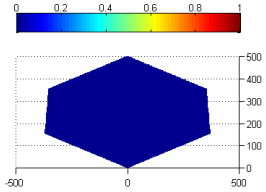
Figure 49. Heat maps for descriptor LS in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



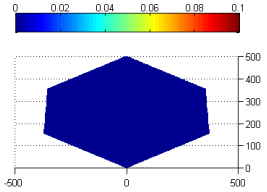
(a) image at 200° rotation
(b) SU precision at 200° rotation
(c) SU recall at 200° rotation
(d) SU efficiency at 200° rotation
(e) image at 325° rotation
(f) SU precision at 325° rotation
(g) SU recall at 325° rotation
(h) SU efficiency at 325° rotation
(i) image at 000° rotation
(j) SU precision at 000° rotation
(k) SU recall at 000° rotation
(l) SU efficiency at 000° rotation
(m) image at 035° rotation
(n) SU precision at 035° rotation
(o) SU recall at 035° rotation
(p) SU efficiency at 035° rotation
(q) image at 160° rotation
(r) SU precision at 160° rotation
(s) SU recall at 160° rotation
(t) SU efficiency at 160° rotation

Figure 50. Heat maps for descriptor SU in the Ballroom scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.

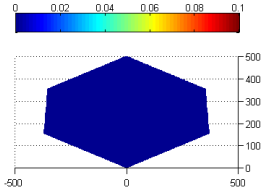
(a) image at 200° rotation



(b) HS precision at 200° rotation

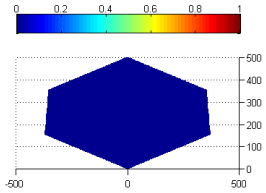


(c) HS recall at 200° rotation

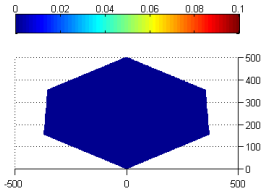


(d) HS efficiency at 200° rotation

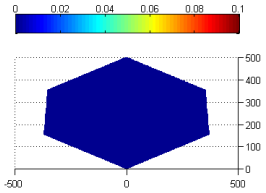
(e) image at 325° rotation



(f) HS precision at 325° rotation

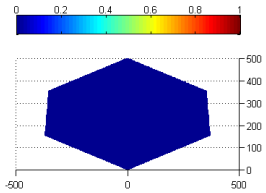


(g) HS recall at 325° rotation

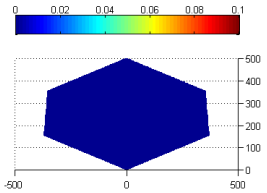


(h) HS efficiency at 325° rotation

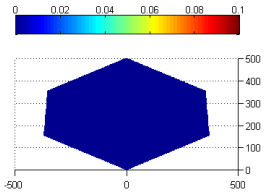
(i) image at 000° rotation



(j) HS precision at 000° rotation

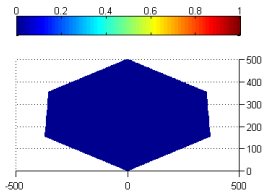


(k) HS recall at 000° rotation

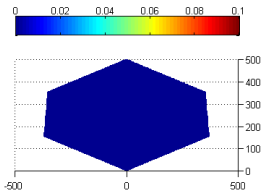


(l) HS efficiency at 000° rotation

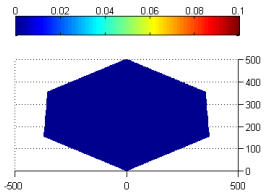
(m) image at 035° rotation



(n) HS precision at 035° rotation

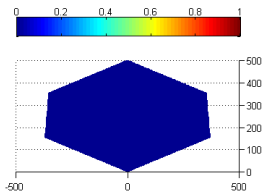


(o) HS recall at 035° rotation

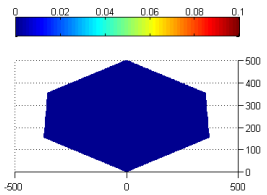


(p) HS efficiency at 035° rotation

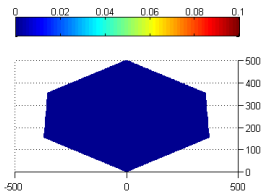
(q) image at 160° rotation



(r) HS precision at 160° rotation

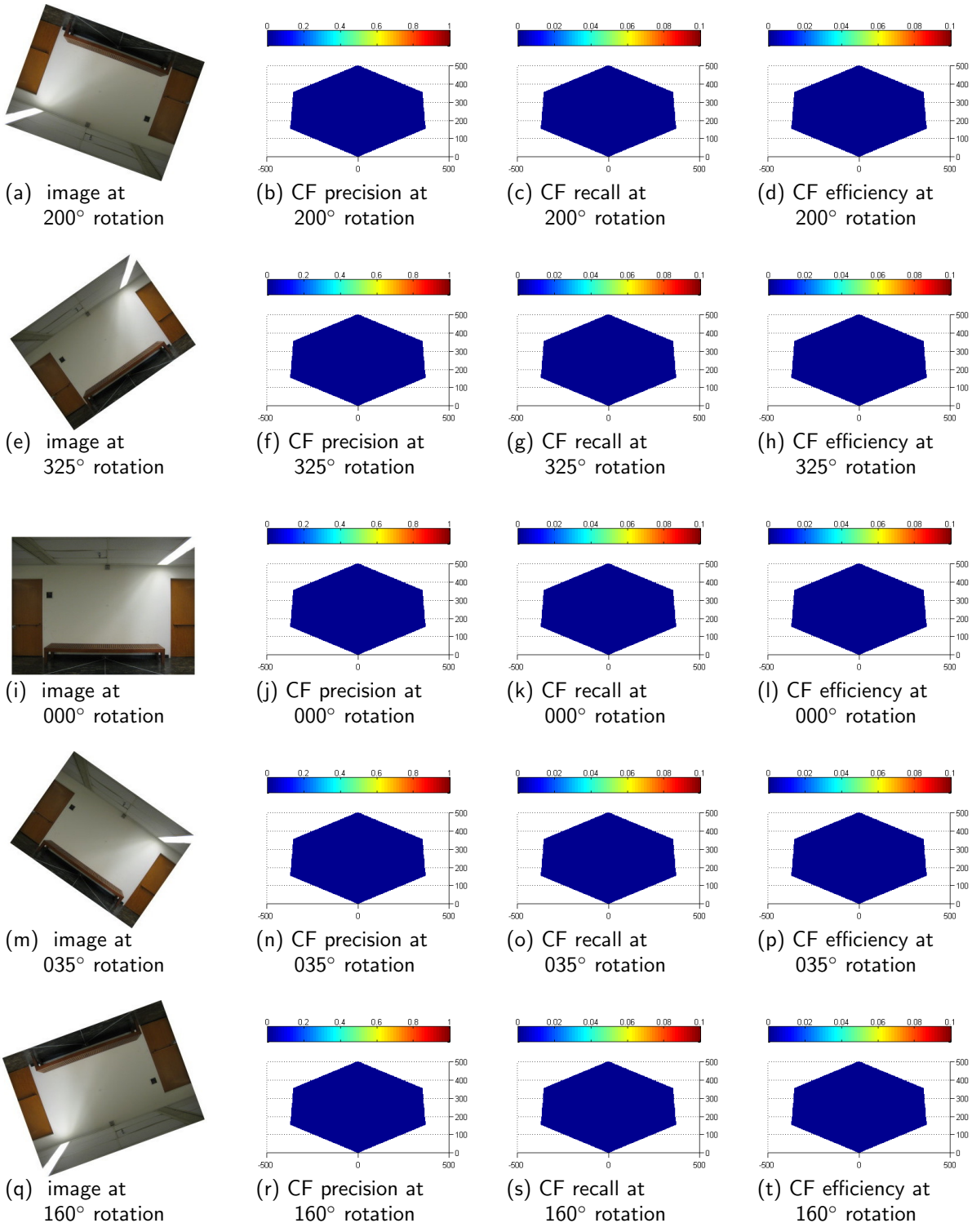


(s) HS recall at 160° rotation



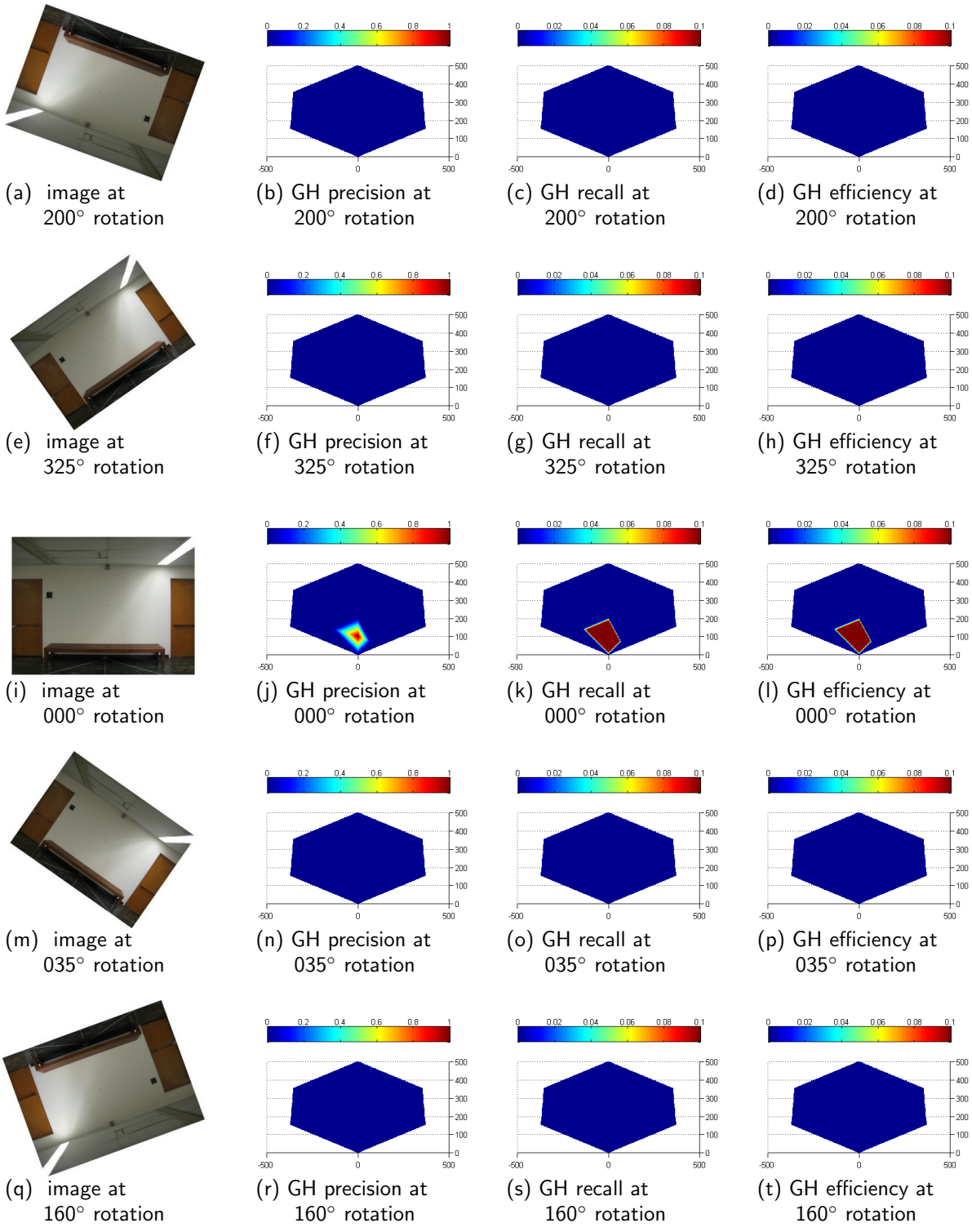
(t) HS efficiency at 160° rotation

Figure 51. Heat maps for descriptor HS in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



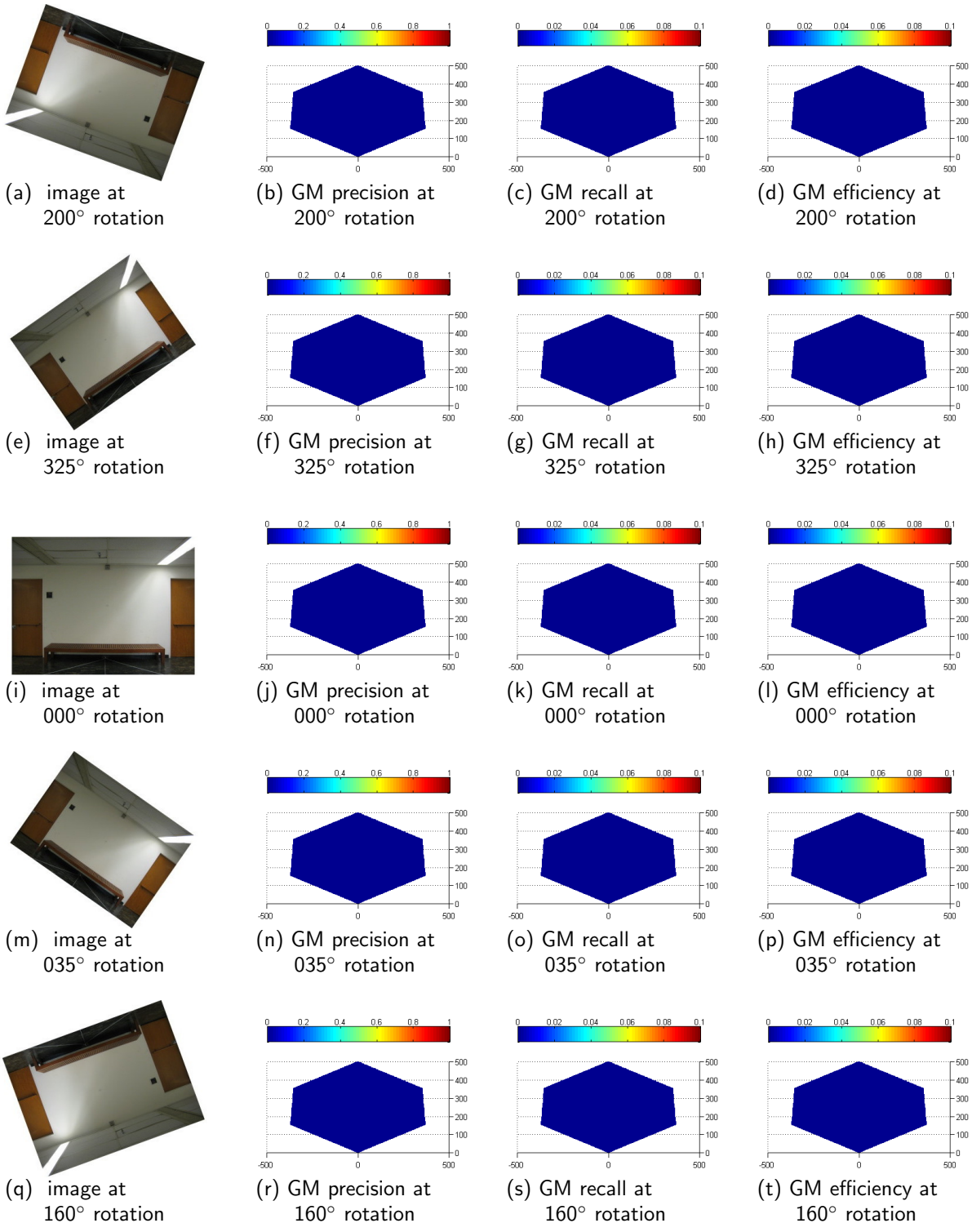
(a) image at 200° rotation
(b) CF precision at 200° rotation
(c) CF recall at 200° rotation
(d) CF efficiency at 200° rotation
(e) image at 325° rotation
(f) CF precision at 325° rotation
(g) CF recall at 325° rotation
(h) CF efficiency at 325° rotation
(i) image at 000° rotation
(j) CF precision at 000° rotation
(k) CF recall at 000° rotation
(l) CF efficiency at 000° rotation
(m) image at 035° rotation
(n) CF precision at 035° rotation
(o) CF recall at 035° rotation
(p) CF efficiency at 035° rotation
(q) image at 160° rotation
(r) CF precision at 160° rotation
(s) CF recall at 160° rotation
(t) CF efficiency at 160° rotation

Figure 52. Heat maps for descriptor CF in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



(a) image at 200° rotation
(b) GH precision at 200° rotation
(c) GH recall at 200° rotation
(d) GH efficiency at 200° rotation
(e) image at 325° rotation
(f) GH precision at 325° rotation
(g) GH recall at 325° rotation
(h) GH efficiency at 325° rotation
(i) image at 000° rotation
(j) GH precision at 000° rotation
(k) GH recall at 000° rotation
(l) GH efficiency at 000° rotation
(m) image at 035° rotation
(n) GH precision at 035° rotation
(o) GH recall at 035° rotation
(p) GH efficiency at 035° rotation
(q) image at 160° rotation
(r) GH precision at 160° rotation
(s) GH recall at 160° rotation
(t) GH efficiency at 160° rotation

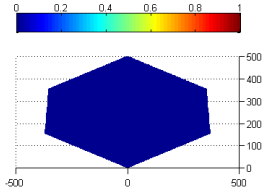
Figure 53. Heat maps for descriptor GH in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



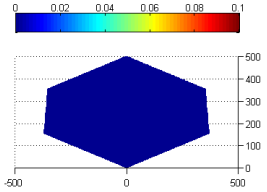
(a) image at 200° rotation
(b) GM precision at 200° rotation
(c) GM recall at 200° rotation
(d) GM efficiency at 200° rotation
(e) image at 325° rotation
(f) GM precision at 325° rotation
(g) GM recall at 325° rotation
(h) GM efficiency at 325° rotation
(i) image at 000° rotation
(j) GM precision at 000° rotation
(k) GM recall at 000° rotation
(l) GM efficiency at 000° rotation
(m) image at 035° rotation
(n) GM precision at 035° rotation
(o) GM recall at 035° rotation
(p) GM efficiency at 035° rotation
(q) image at 160° rotation
(r) GM precision at 160° rotation
(s) GM recall at 160° rotation
(t) GM efficiency at 160° rotation

Figure 54. Heat maps for descriptor GM in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.

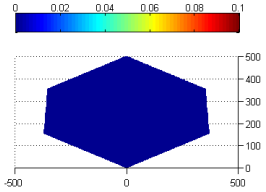
(a) image at 200° rotation



(b) DI precision at 200° rotation

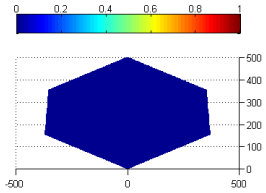


(c) DI recall at 200° rotation

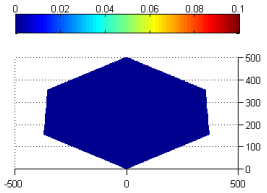


(d) DI efficiency at 200° rotation

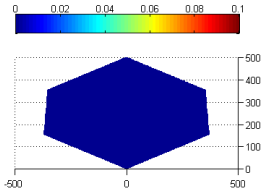
(e) image at 325° rotation



(f) DI precision at 325° rotation

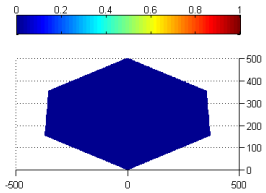


(g) DI recall at 325° rotation

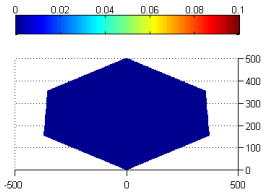


(h) DI efficiency at 325° rotation

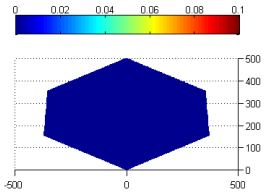
(i) image at 000° rotation



(j) DI precision at 000° rotation

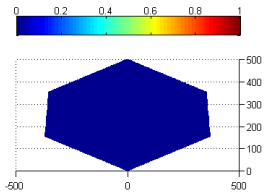


(k) DI recall at 000° rotation

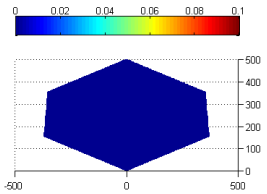


(l) DI efficiency at 000° rotation

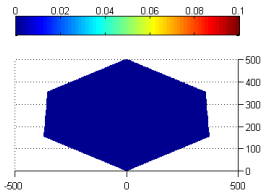
(m) image at 035° rotation



(n) DI precision at 035° rotation

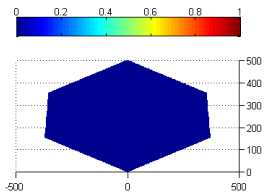


(o) DI recall at 035° rotation

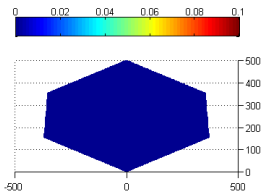


(p) DI efficiency at 035° rotation

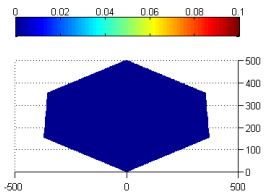
(q) image at 160° rotation



(r) DI precision at 160° rotation

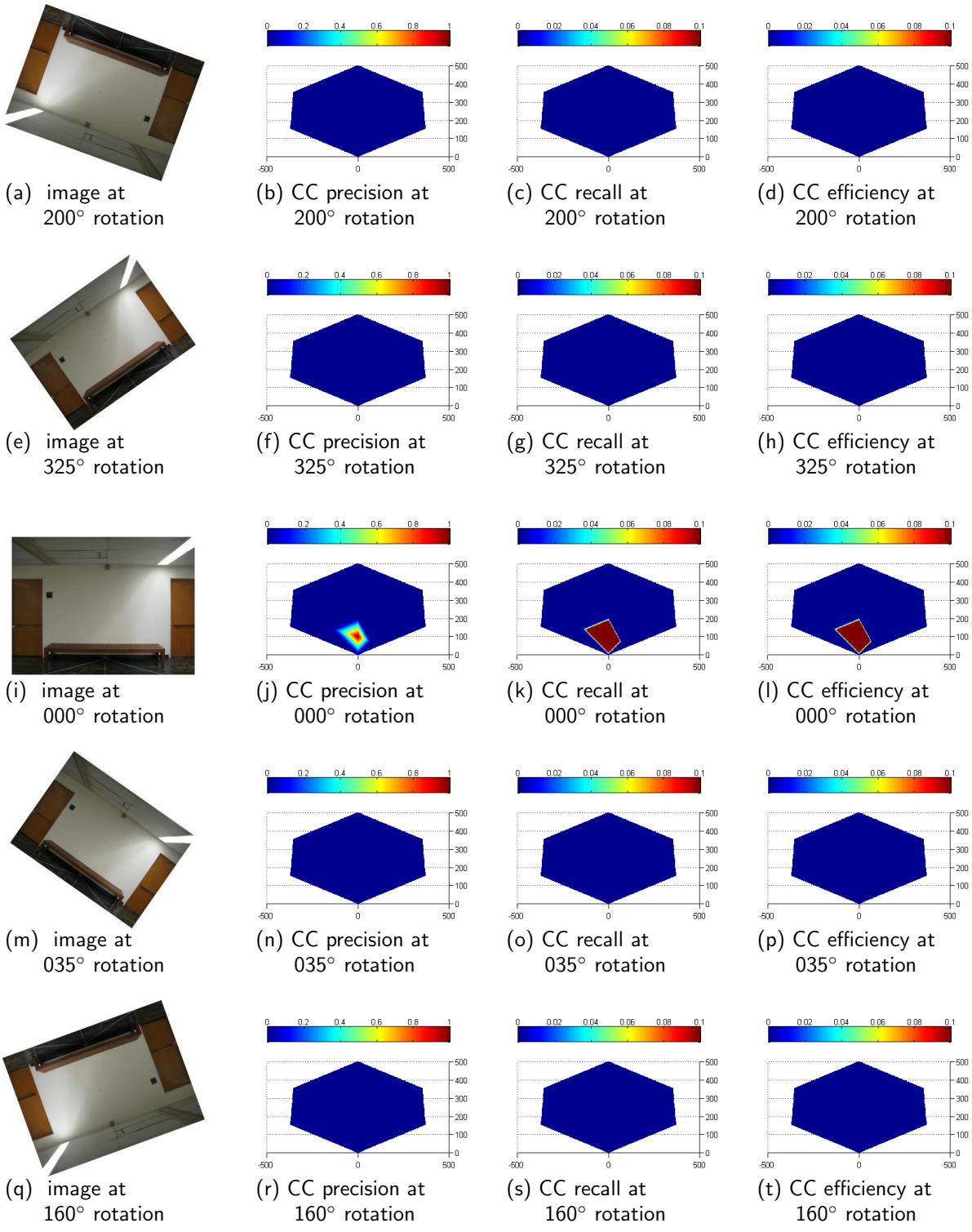


(s) DI recall at 160° rotation



(t) DI efficiency at 160° rotation

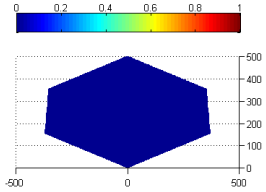
Figure 55. Heat maps for descriptor DI in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



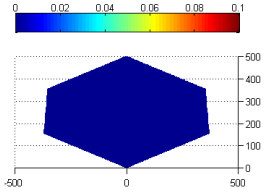
(a) image at 200° rotation
(b) CC precision at 200° rotation
(c) CC recall at 200° rotation
(d) CC efficiency at 200° rotation
(e) image at 325° rotation
(f) CC precision at 325° rotation
(g) CC recall at 325° rotation
(h) CC efficiency at 325° rotation
(i) image at 000° rotation
(j) CC precision at 000° rotation
(k) CC recall at 000° rotation
(l) CC efficiency at 000° rotation
(m) image at 035° rotation
(n) CC precision at 035° rotation
(o) CC recall at 035° rotation
(p) CC efficiency at 035° rotation
(q) image at 160° rotation
(r) CC precision at 160° rotation
(s) CC recall at 160° rotation
(t) CC efficiency at 160° rotation

Figure 56. Heat maps for descriptor CC in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.

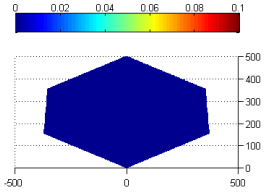
(a) image at 200° rotation



(b) SF precision at 200° rotation

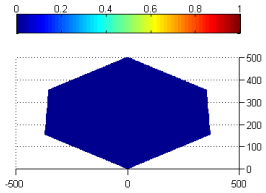


(c) SF recall at 200° rotation

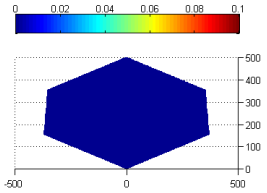


(d) SF efficiency at 200° rotation

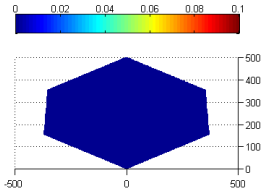
(e) image at 325° rotation



(f) SF precision at 325° rotation

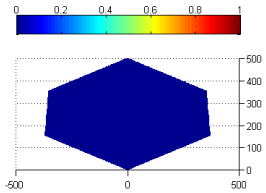


(g) SF recall at 325° rotation

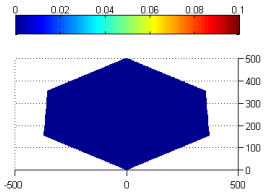


(h) SF efficiency at 325° rotation

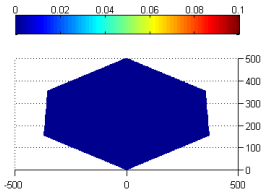
(i) image at 000° rotation



(j) SF precision at 000° rotation

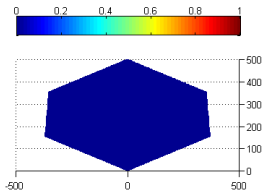


(k) SF recall at 000° rotation

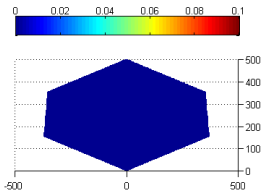


(l) SF efficiency at 000° rotation

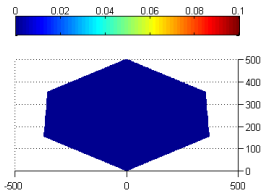
(m) image at 035° rotation



(n) SF precision at 035° rotation

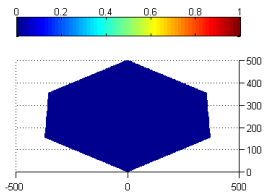


(o) SF recall at 035° rotation

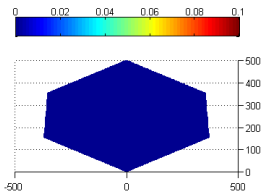


(p) SF efficiency at 035° rotation

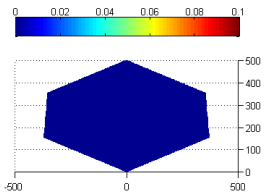
(q) image at 160° rotation



(r) SF precision at 160° rotation



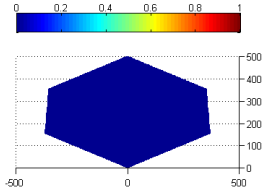
(s) SF recall at 160° rotation



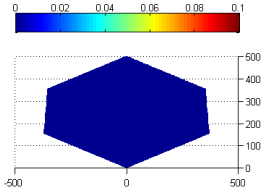
(t) SF efficiency at 160° rotation

Figure 57. Heat maps for descriptor SF in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.

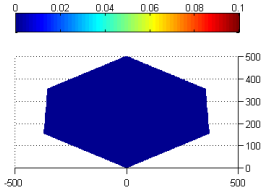
(a) image at 200° rotation



(b) PS precision at 200° rotation

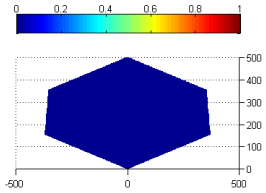


(c) PS recall at 200° rotation

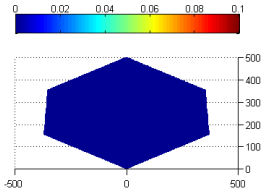


(d) PS efficiency at 200° rotation

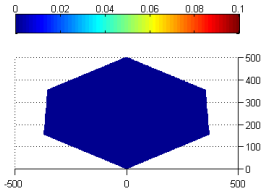
(e) image at 325° rotation



(f) PS precision at 325° rotation

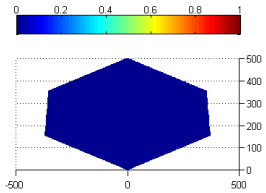


(g) PS recall at 325° rotation

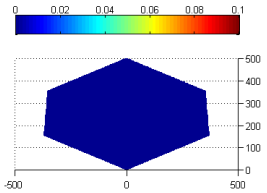


(h) PS efficiency at 325° rotation

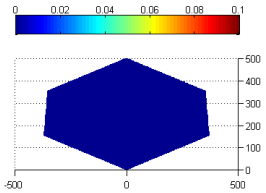
(i) image at 000° rotation



(j) PS precision at 000° rotation

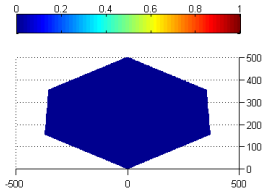


(k) PS recall at 000° rotation

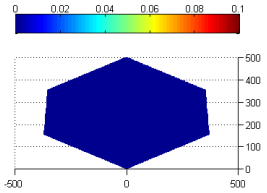


(l) PS efficiency at 000° rotation

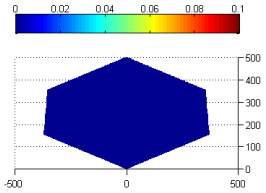
(m) image at 035° rotation



(n) PS precision at 035° rotation

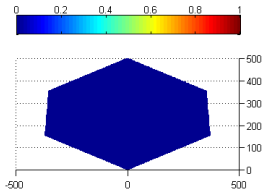


(o) PS recall at 035° rotation

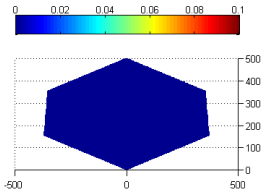


(p) PS efficiency at 035° rotation

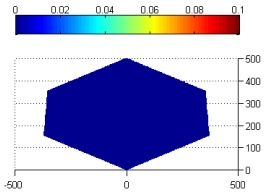
(q) image at 160° rotation



(r) PS precision at 160° rotation

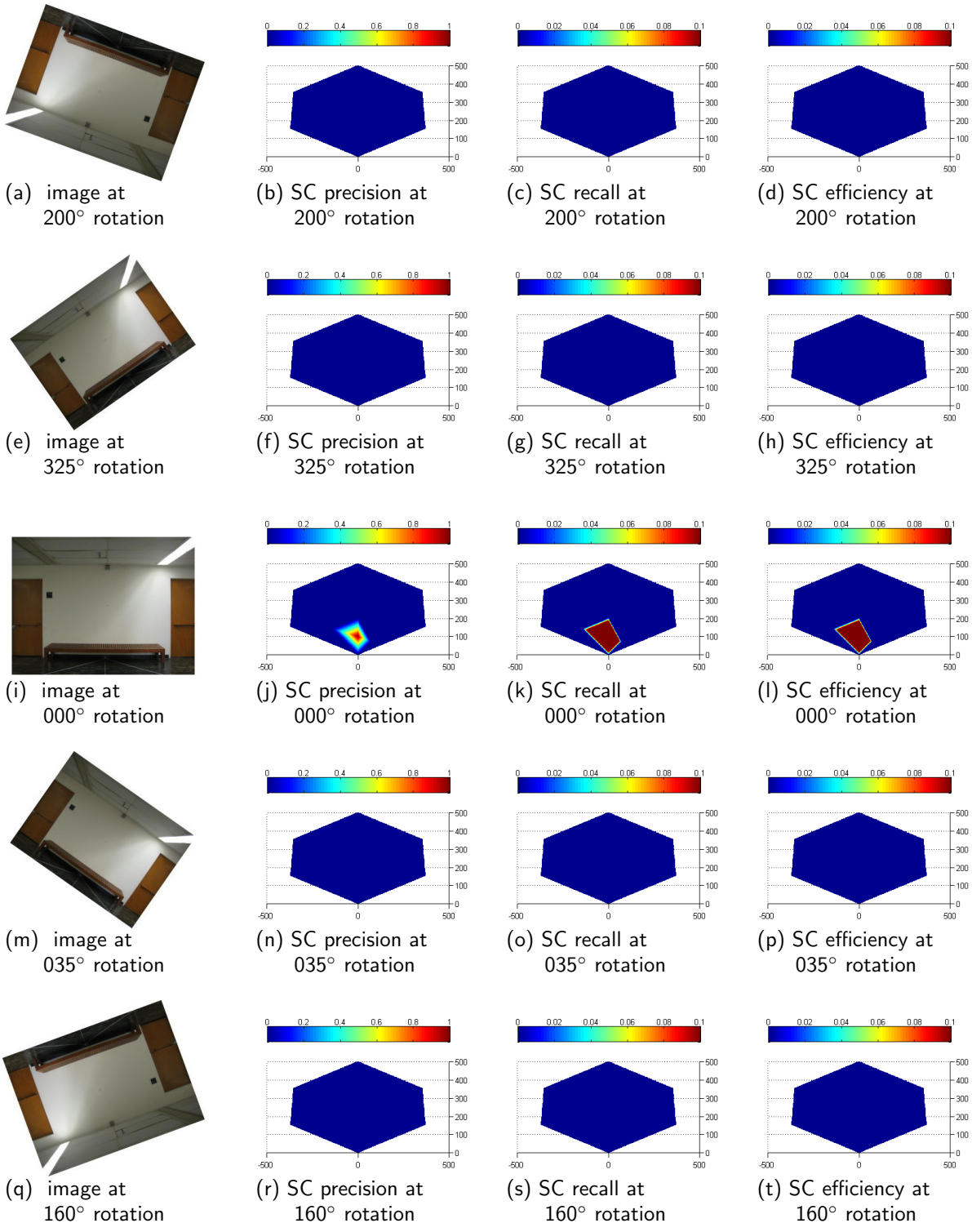


(s) PS recall at 160° rotation



(t) PS efficiency at 160° rotation

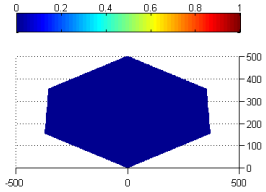
Figure 58. Heat maps for descriptor PS in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



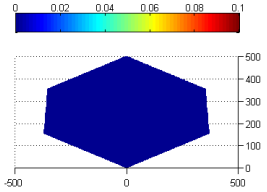
(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

Figure 59. Heat maps for descriptor SC in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.

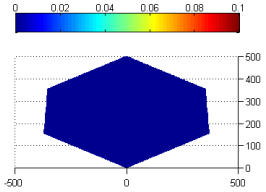
(a) image at 200° rotation



(b) LS precision at 200° rotation

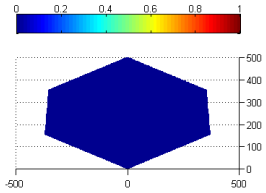


(c) LS recall at 200° rotation

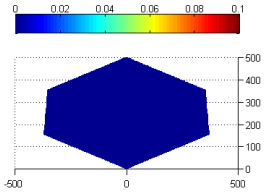


(d) LS efficiency at 200° rotation

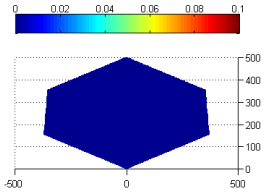
(e) image at 325° rotation



(f) LS precision at 325° rotation

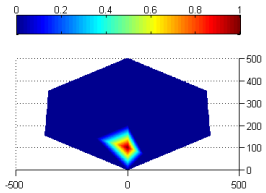


(g) LS recall at 325° rotation

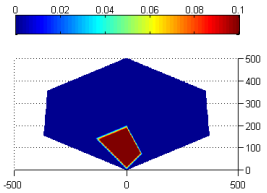


(h) LS efficiency at 325° rotation

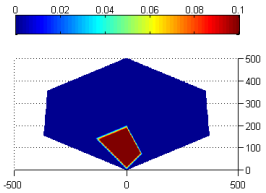
(i) image at 000° rotation



(j) LS precision at 000° rotation

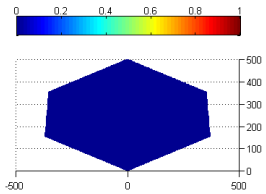


(k) LS recall at 000° rotation

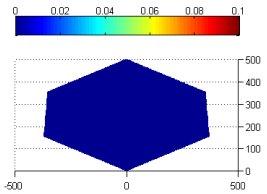


(l) LS efficiency at 000° rotation

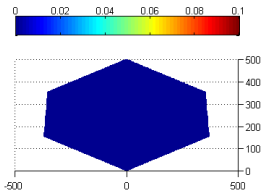
(m) image at 035° rotation



(n) LS precision at 035° rotation

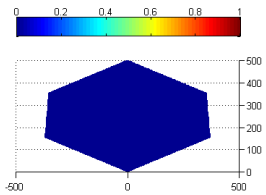


(o) LS recall at 035° rotation

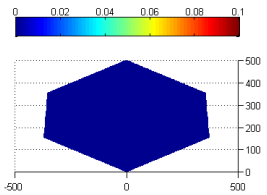


(p) LS efficiency at 035° rotation

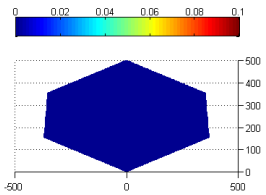
(q) image at 160° rotation



(r) LS precision at 160° rotation

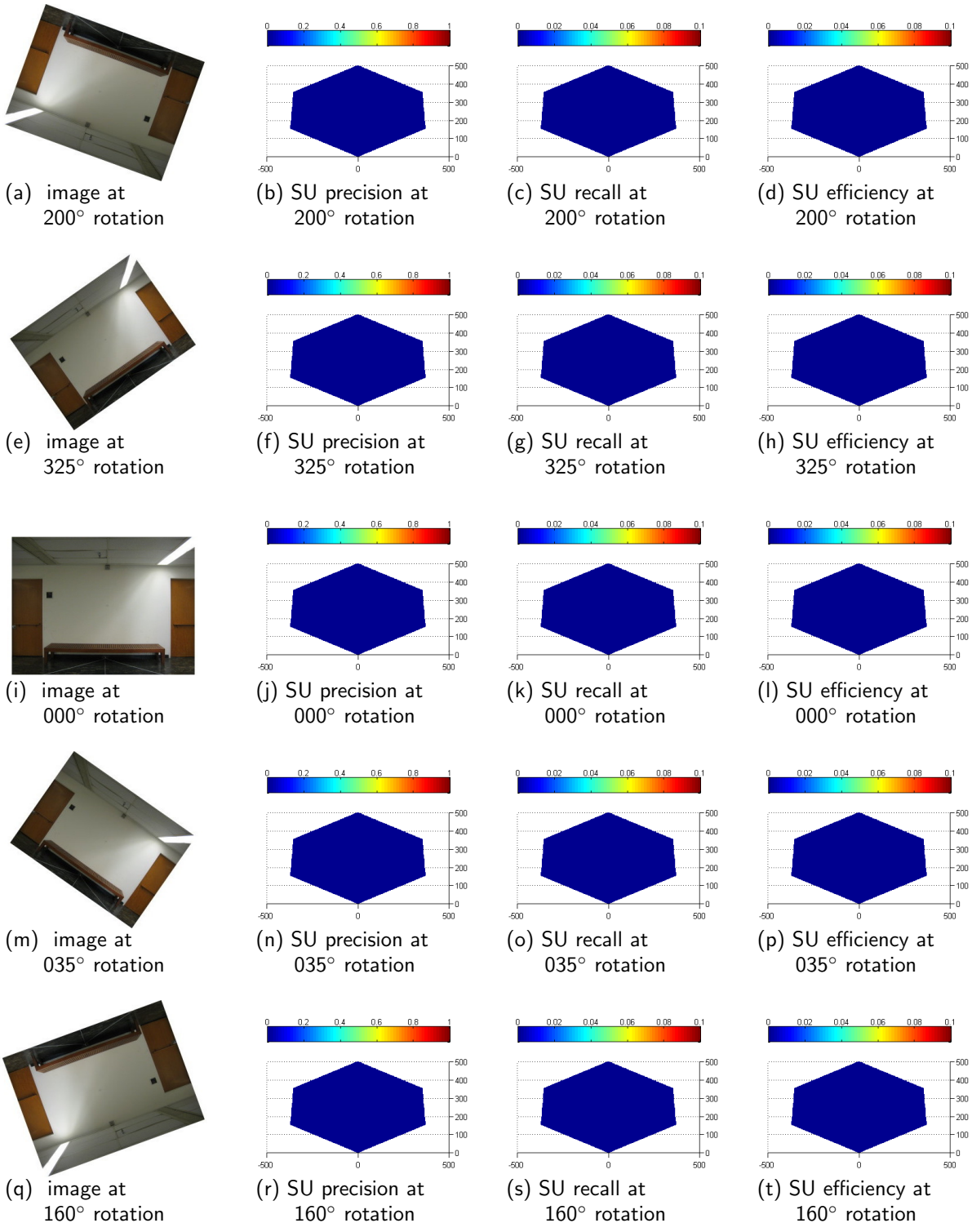


(s) LS recall at 160° rotation



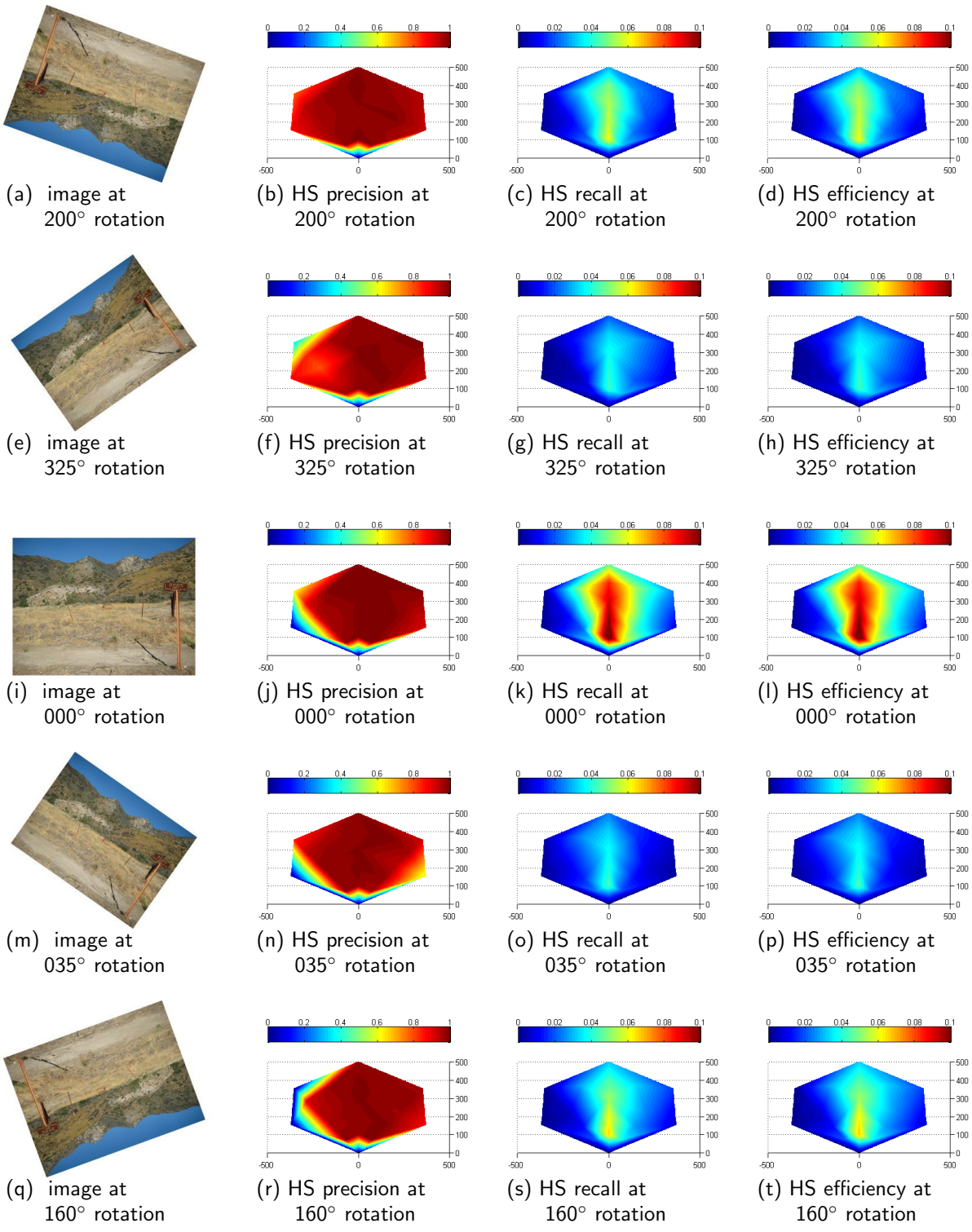
(t) LS efficiency at 160° rotation

Figure 60. Heat maps for descriptor LS in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



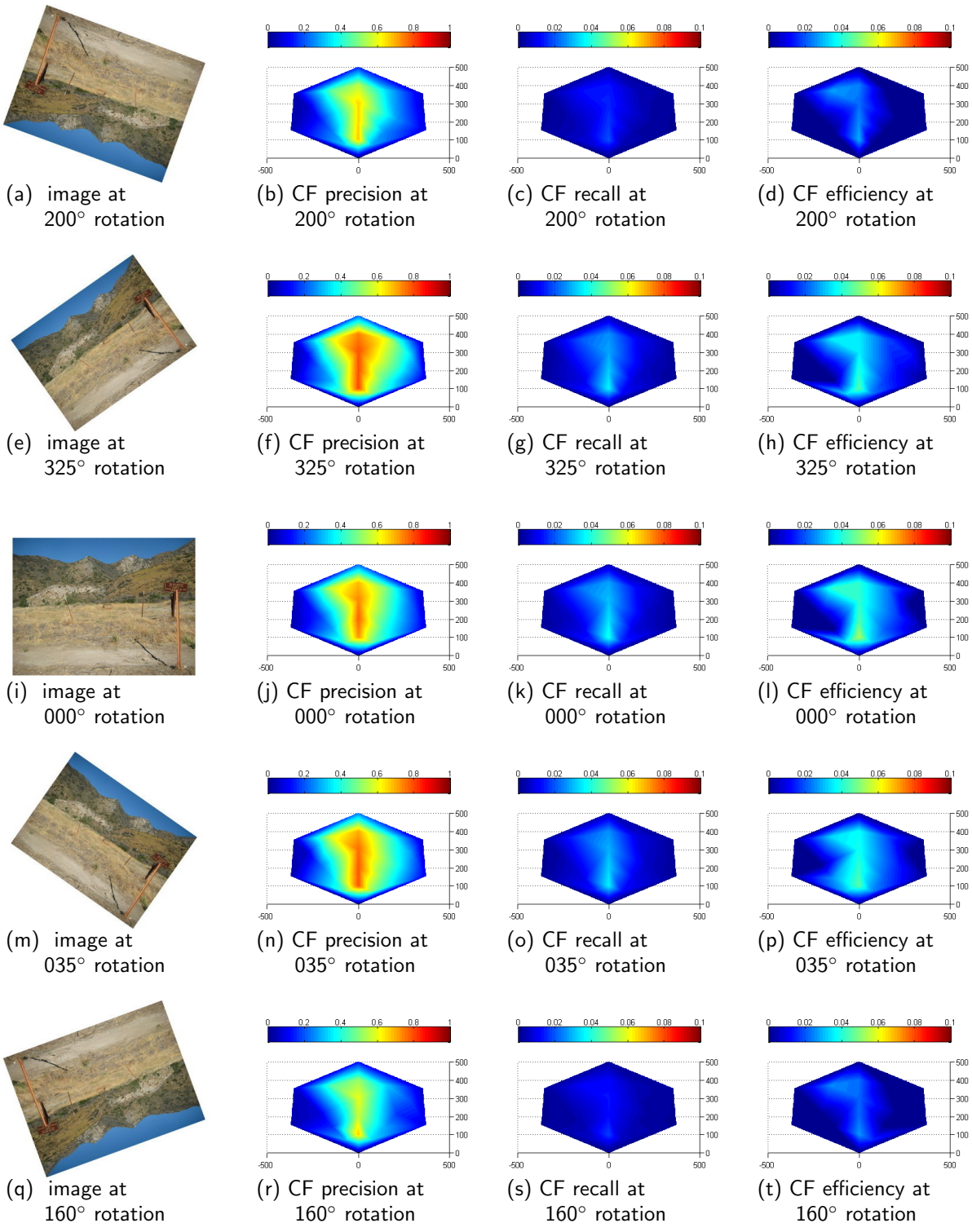
(a) image at 200° rotation
(b) SU precision at 200° rotation
(c) SU recall at 200° rotation
(d) SU efficiency at 200° rotation
(e) image at 325° rotation
(f) SU precision at 325° rotation
(g) SU recall at 325° rotation
(h) SU efficiency at 325° rotation
(i) image at 000° rotation
(j) SU precision at 000° rotation
(k) SU recall at 000° rotation
(l) SU efficiency at 000° rotation
(m) image at 035° rotation
(n) SU precision at 035° rotation
(o) SU recall at 035° rotation
(p) SU efficiency at 035° rotation
(q) image at 160° rotation
(r) SU precision at 160° rotation
(s) SU recall at 160° rotation
(t) SU efficiency at 160° rotation

Figure 61. Heat maps for descriptor SU in the KingHall scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=1m.



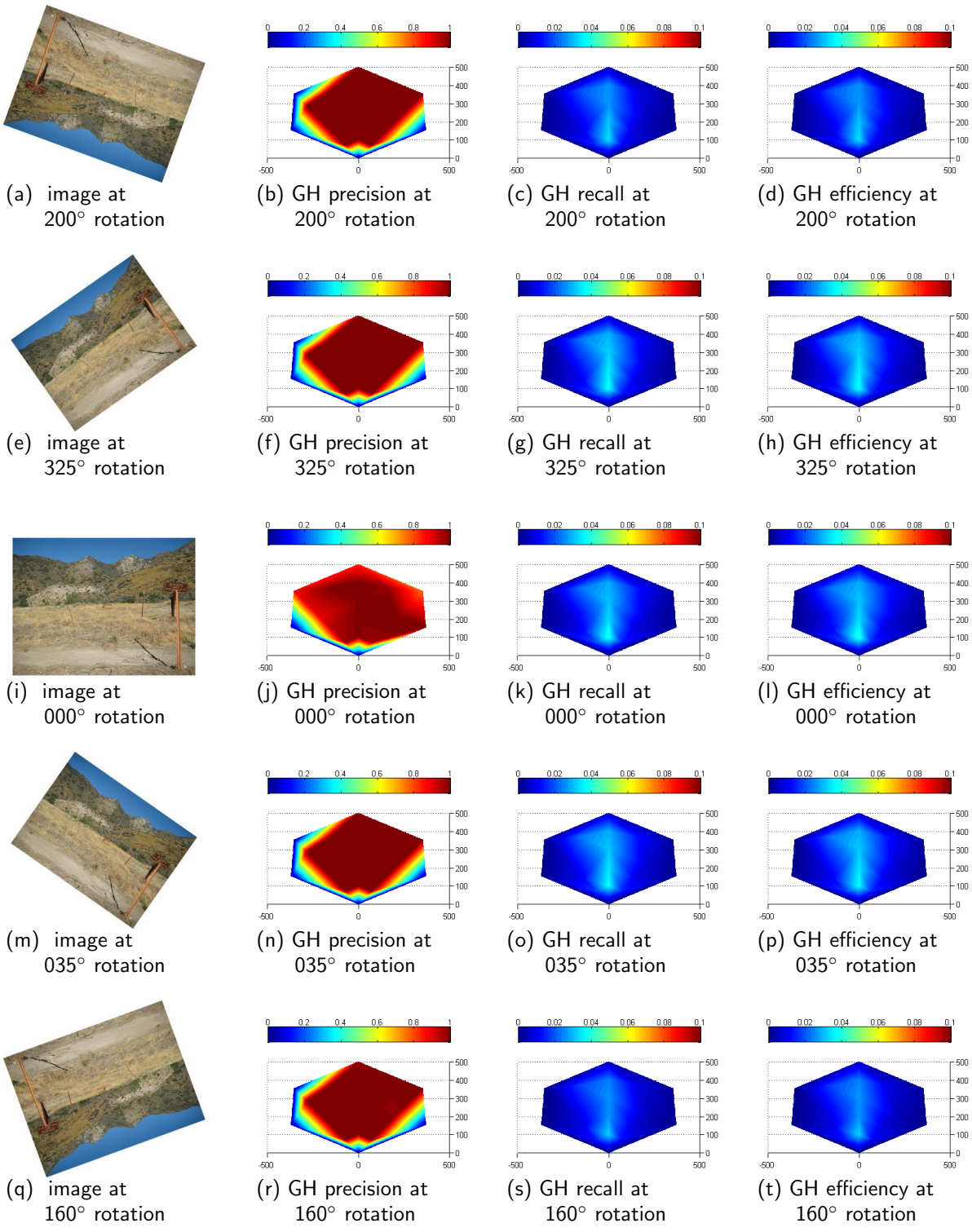
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 62. Heat maps for descriptor HS in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



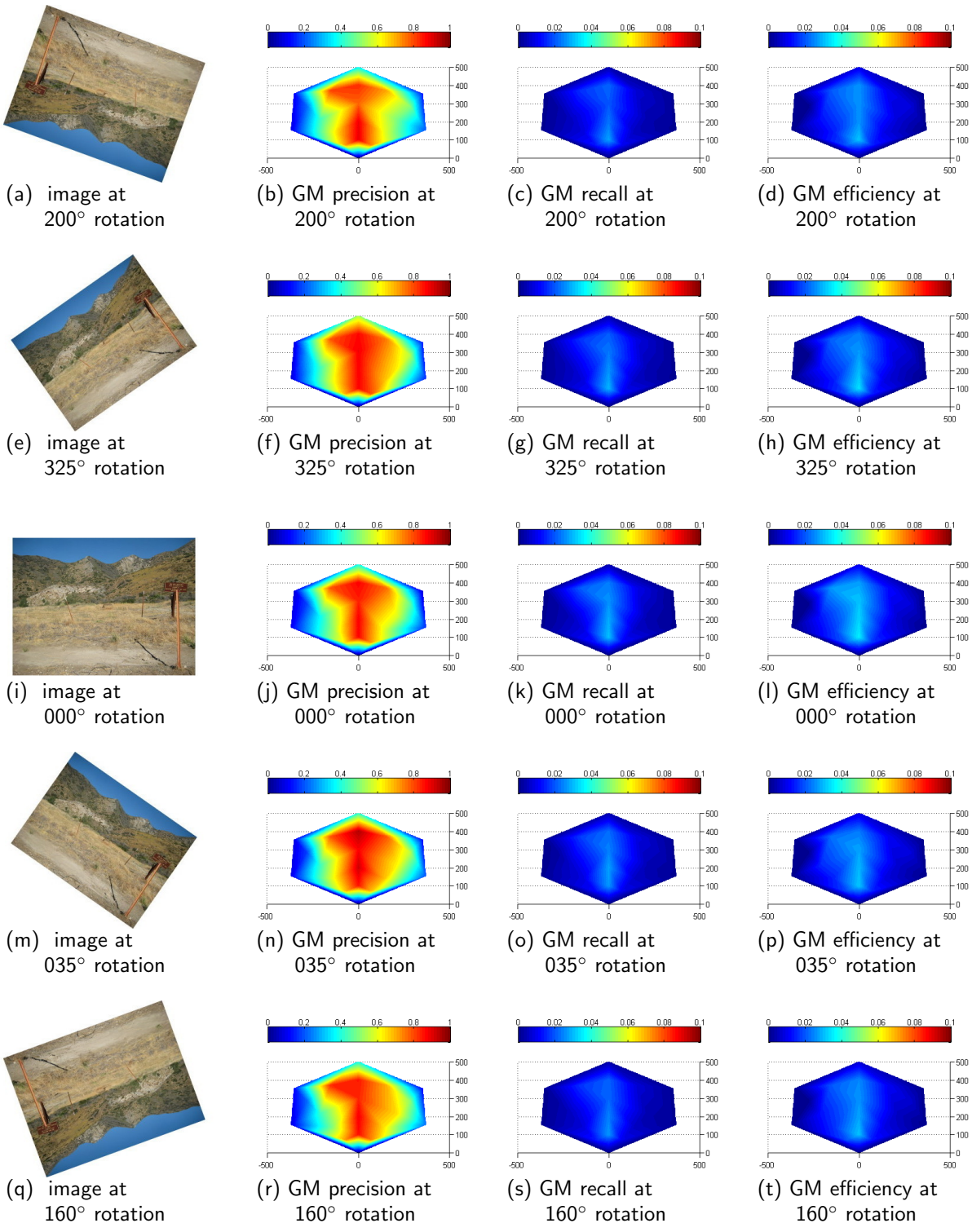
(a) image at 200° rotation
(b) CF precision at 200° rotation
(c) CF recall at 200° rotation
(d) CF efficiency at 200° rotation
(e) image at 325° rotation
(f) CF precision at 325° rotation
(g) CF recall at 325° rotation
(h) CF efficiency at 325° rotation
(i) image at 000° rotation
(j) CF precision at 000° rotation
(k) CF recall at 000° rotation
(l) CF efficiency at 000° rotation
(m) image at 035° rotation
(n) CF precision at 035° rotation
(o) CF recall at 035° rotation
(p) CF efficiency at 035° rotation
(q) image at 160° rotation
(r) CF precision at 160° rotation
(s) CF recall at 160° rotation
(t) CF efficiency at 160° rotation

Figure 63. Heat maps for descriptor CF in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



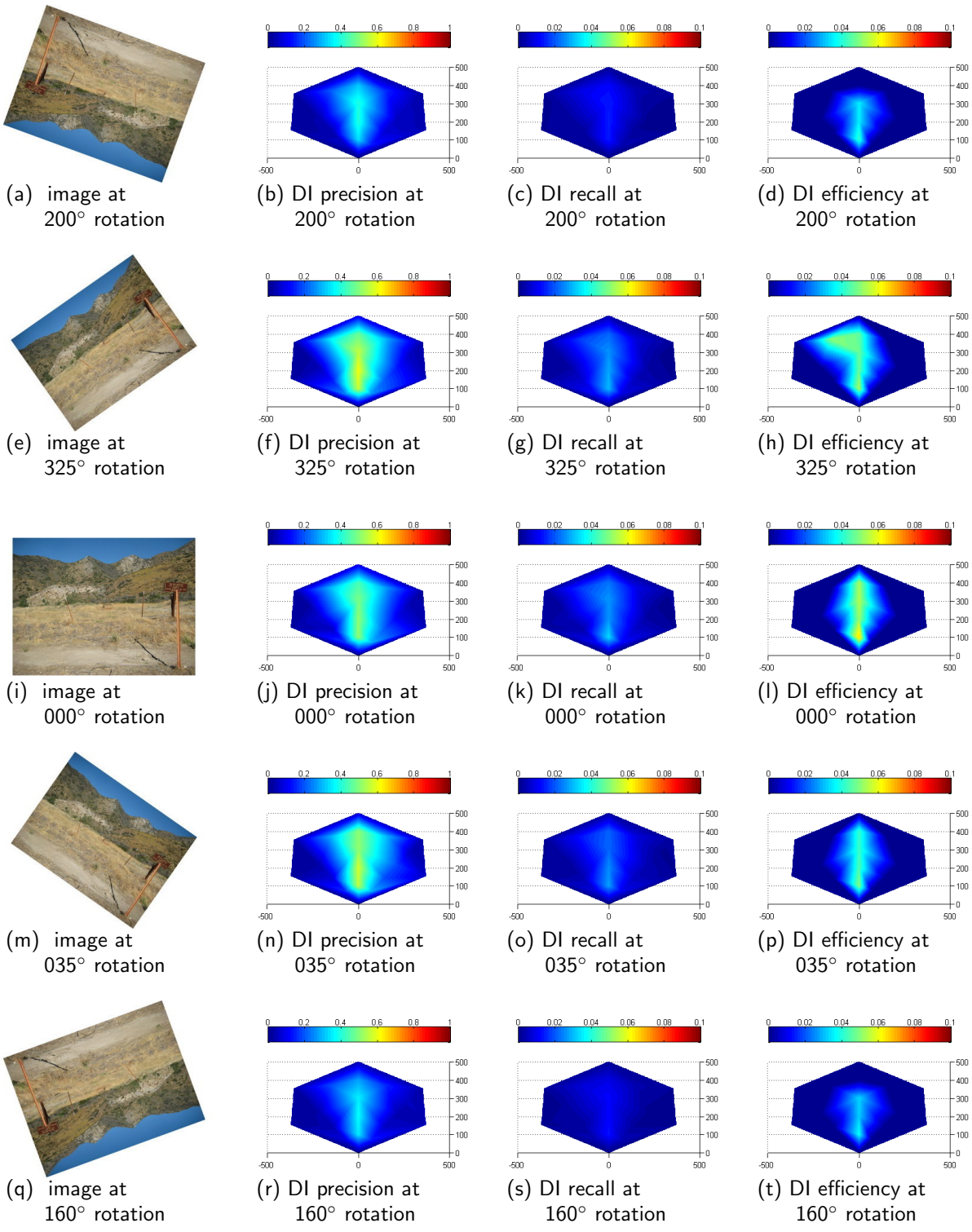
(a) image at 200° rotation
(b) GH precision at 200° rotation
(c) GH recall at 200° rotation
(d) GH efficiency at 200° rotation
(e) image at 325° rotation
(f) GH precision at 325° rotation
(g) GH recall at 325° rotation
(h) GH efficiency at 325° rotation
(i) image at 000° rotation
(j) GH precision at 000° rotation
(k) GH recall at 000° rotation
(l) GH efficiency at 000° rotation
(m) image at 035° rotation
(n) GH precision at 035° rotation
(o) GH recall at 035° rotation
(p) GH efficiency at 035° rotation
(q) image at 160° rotation
(r) GH precision at 160° rotation
(s) GH recall at 160° rotation
(t) GH efficiency at 160° rotation

Figure 64. Heat maps for descriptor GH in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



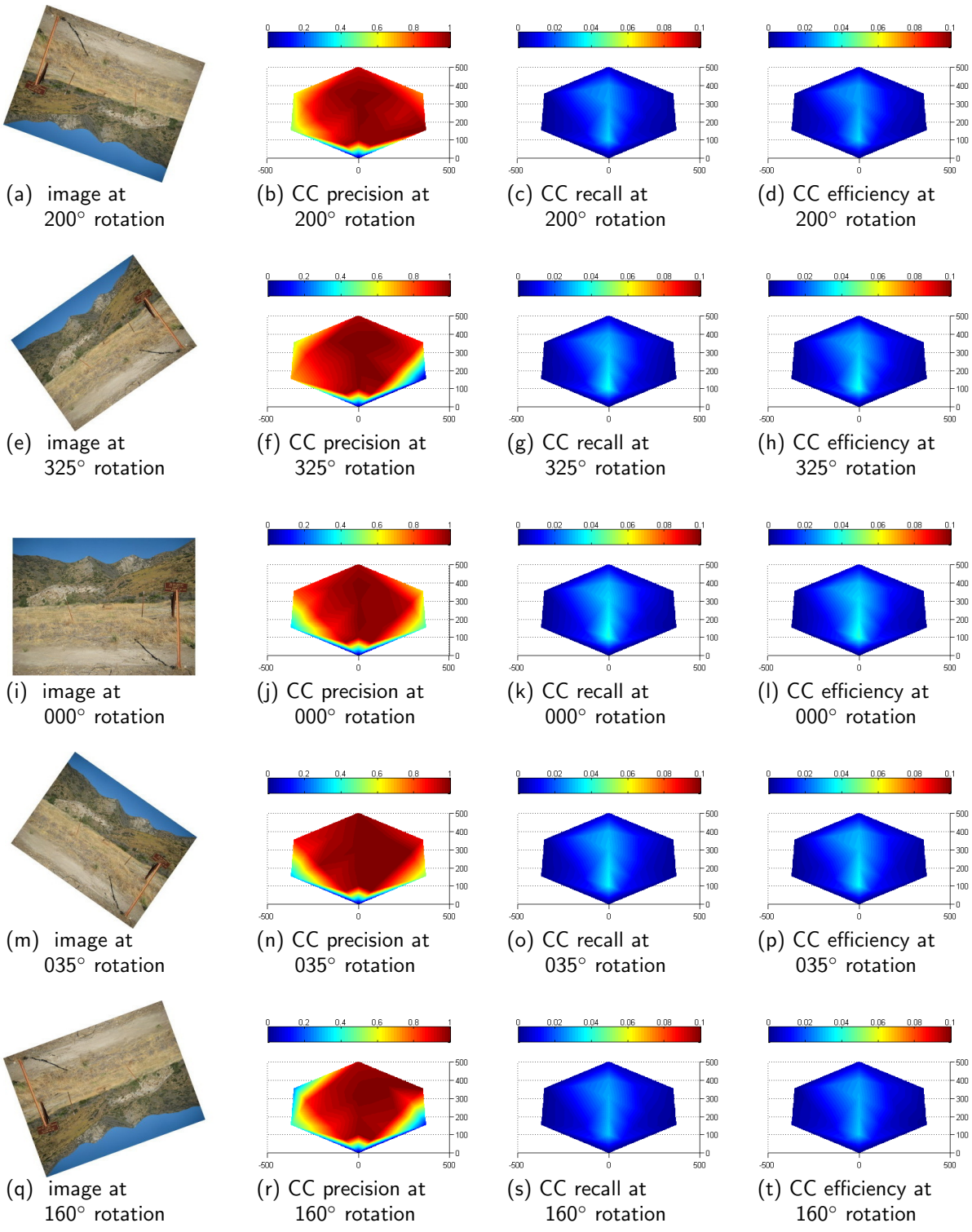
(a) image at 200° rotation
(b) GM precision at 200° rotation
(c) GM recall at 200° rotation
(d) GM efficiency at 200° rotation
(e) image at 325° rotation
(f) GM precision at 325° rotation
(g) GM recall at 325° rotation
(h) GM efficiency at 325° rotation
(i) image at 000° rotation
(j) GM precision at 000° rotation
(k) GM recall at 000° rotation
(l) GM efficiency at 000° rotation
(m) image at 035° rotation
(n) GM precision at 035° rotation
(o) GM recall at 035° rotation
(p) GM efficiency at 035° rotation
(q) image at 160° rotation
(r) GM precision at 160° rotation
(s) GM recall at 160° rotation
(t) GM efficiency at 160° rotation

Figure 65. Heat maps for descriptor GM in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



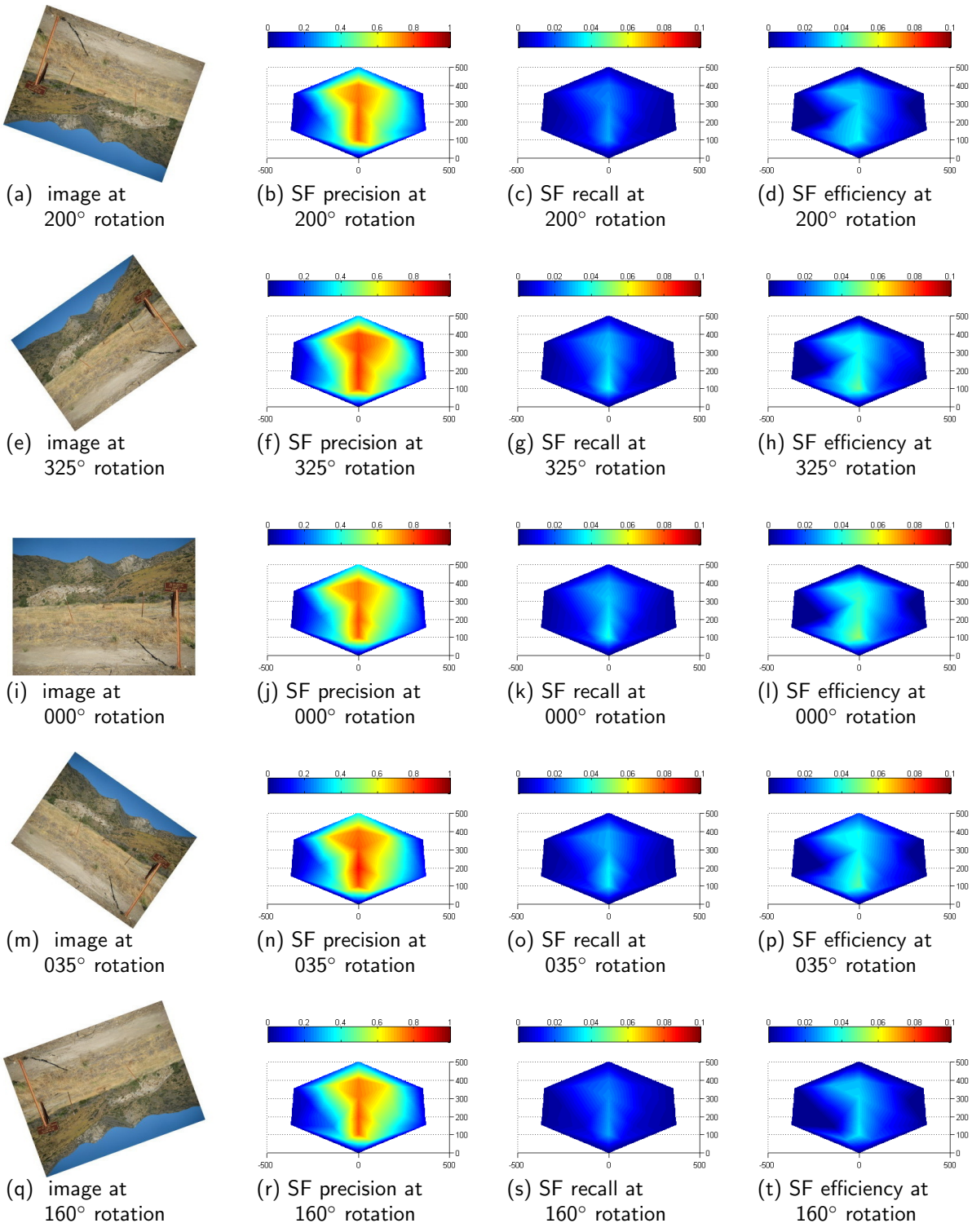
(a) image at 200° rotation
(b) DI precision at 200° rotation
(c) DI recall at 200° rotation
(d) DI efficiency at 200° rotation
(e) image at 325° rotation
(f) DI precision at 325° rotation
(g) DI recall at 325° rotation
(h) DI efficiency at 325° rotation
(i) image at 000° rotation
(j) DI precision at 000° rotation
(k) DI recall at 000° rotation
(l) DI efficiency at 000° rotation
(m) image at 035° rotation
(n) DI precision at 035° rotation
(o) DI recall at 035° rotation
(p) DI efficiency at 035° rotation
(q) image at 160° rotation
(r) DI precision at 160° rotation
(s) DI recall at 160° rotation
(t) DI efficiency at 160° rotation

Figure 66. Heat maps for descriptor DI in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



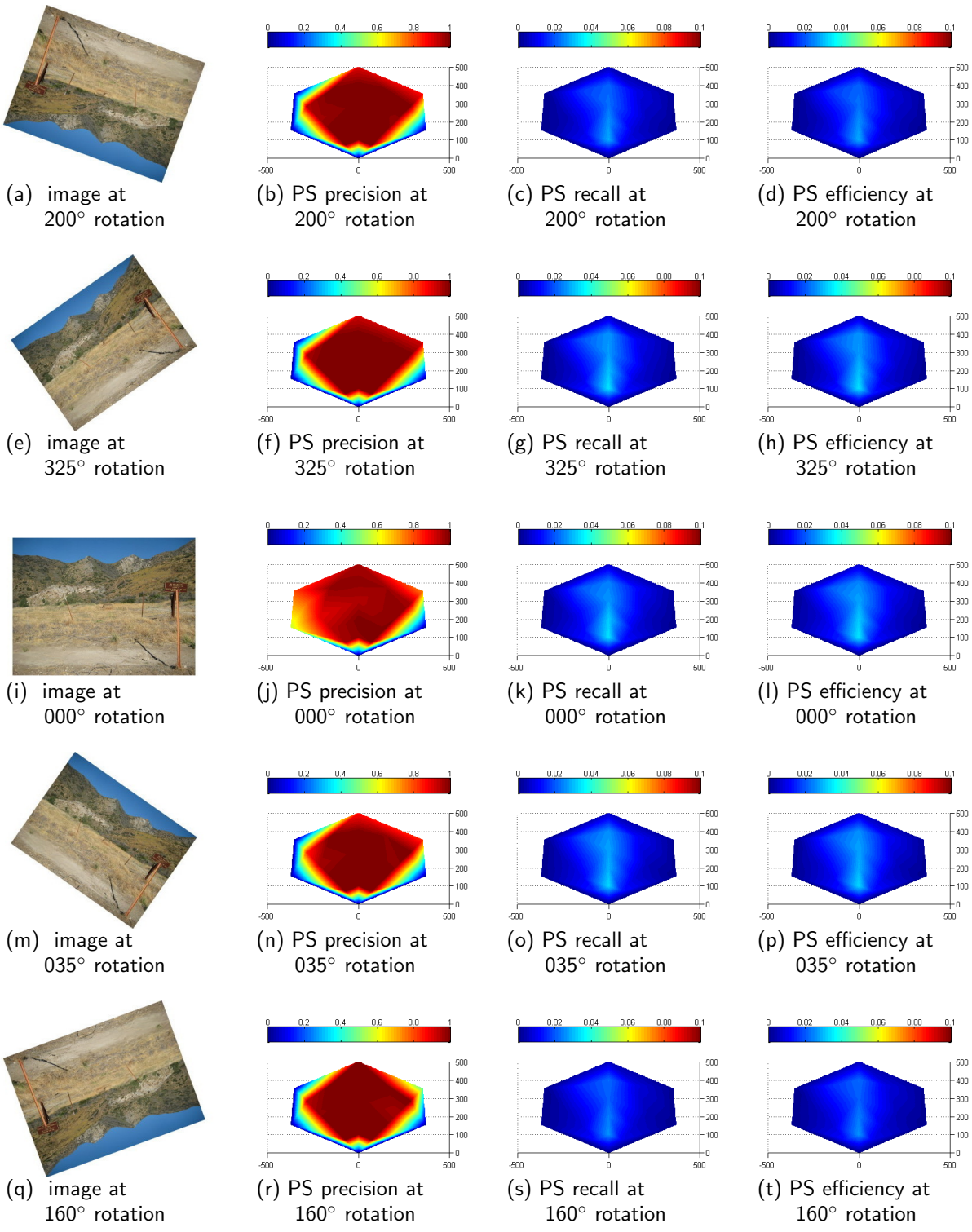
(a) image at 200° rotation
(b) CC precision at 200° rotation
(c) CC recall at 200° rotation
(d) CC efficiency at 200° rotation
(e) image at 325° rotation
(f) CC precision at 325° rotation
(g) CC recall at 325° rotation
(h) CC efficiency at 325° rotation
(i) image at 000° rotation
(j) CC precision at 000° rotation
(k) CC recall at 000° rotation
(l) CC efficiency at 000° rotation
(m) image at 035° rotation
(n) CC precision at 035° rotation
(o) CC recall at 035° rotation
(p) CC efficiency at 035° rotation
(q) image at 160° rotation
(r) CC precision at 160° rotation
(s) CC recall at 160° rotation
(t) CC efficiency at 160° rotation

Figure 67. Heat maps for descriptor CC in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



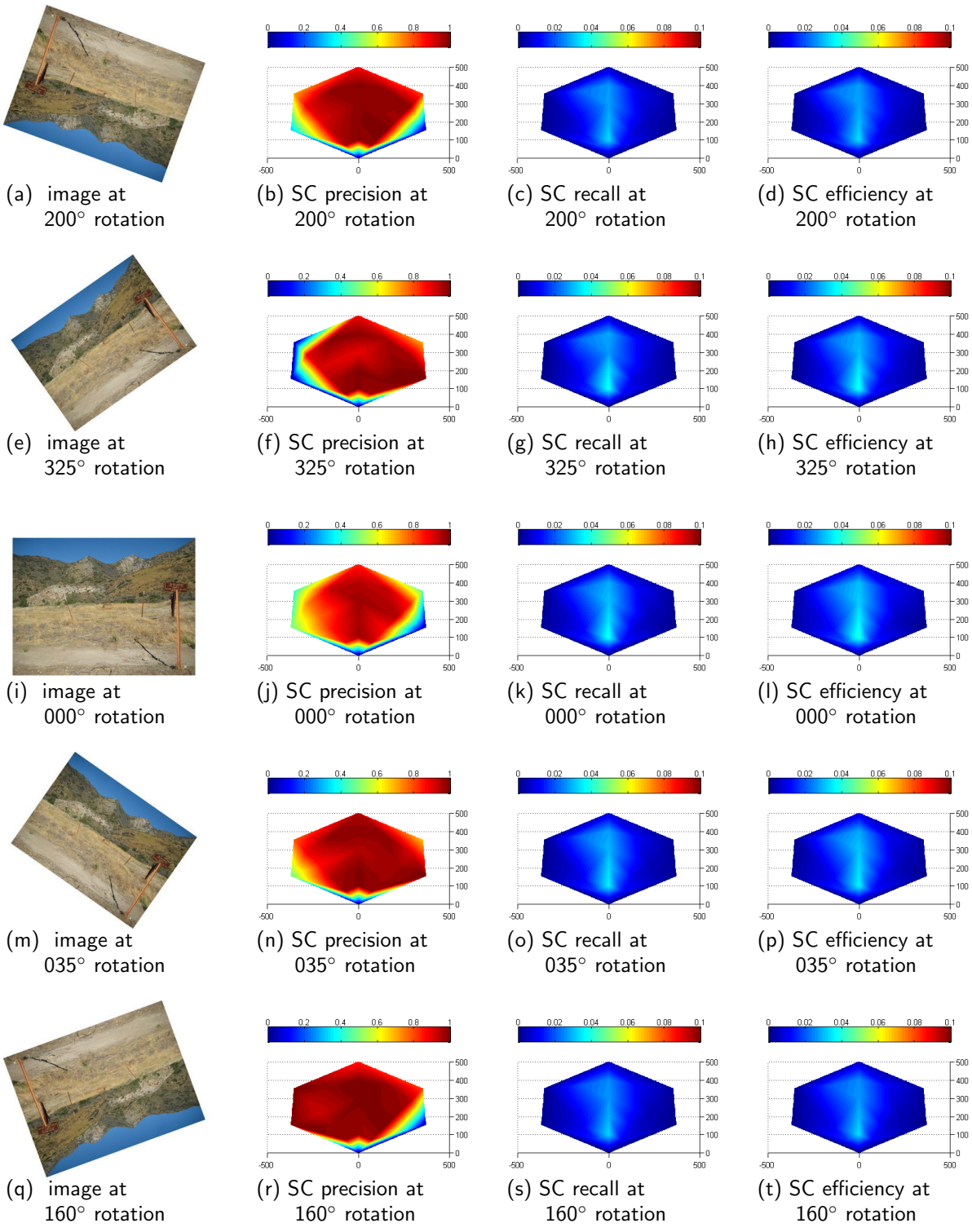
(a) image at 200° rotation
(b) SF precision at 200° rotation
(c) SF recall at 200° rotation
(d) SF efficiency at 200° rotation
(e) image at 325° rotation
(f) SF precision at 325° rotation
(g) SF recall at 325° rotation
(h) SF efficiency at 325° rotation
(i) image at 000° rotation
(j) SF precision at 000° rotation
(k) SF recall at 000° rotation
(l) SF efficiency at 000° rotation
(m) image at 035° rotation
(n) SF precision at 035° rotation
(o) SF recall at 035° rotation
(p) SF efficiency at 035° rotation
(q) image at 160° rotation
(r) SF precision at 160° rotation
(s) SF recall at 160° rotation
(t) SF efficiency at 160° rotation

Figure 68. Heat maps for descriptor SF in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



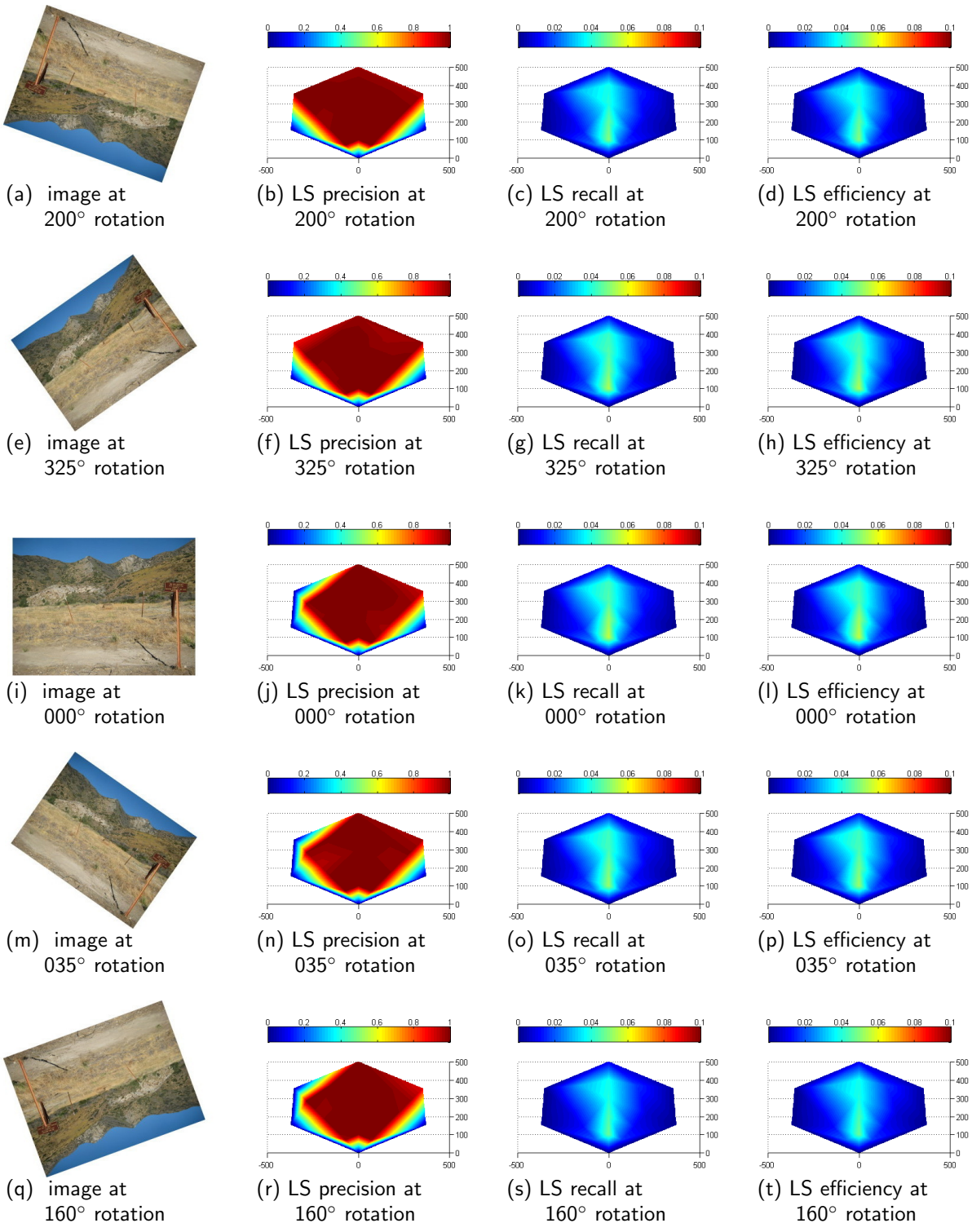
(a) image at 200° rotation
(b) PS precision at 200° rotation
(c) PS recall at 200° rotation
(d) PS efficiency at 200° rotation
(e) image at 325° rotation
(f) PS precision at 325° rotation
(g) PS recall at 325° rotation
(h) PS efficiency at 325° rotation
(i) image at 000° rotation
(j) PS precision at 000° rotation
(k) PS recall at 000° rotation
(l) PS efficiency at 000° rotation
(m) image at 035° rotation
(n) PS precision at 035° rotation
(o) PS recall at 035° rotation
(p) PS efficiency at 035° rotation
(q) image at 160° rotation
(r) PS precision at 160° rotation
(s) PS recall at 160° rotation
(t) PS efficiency at 160° rotation

Figure 69. Heat maps for descriptor PS in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



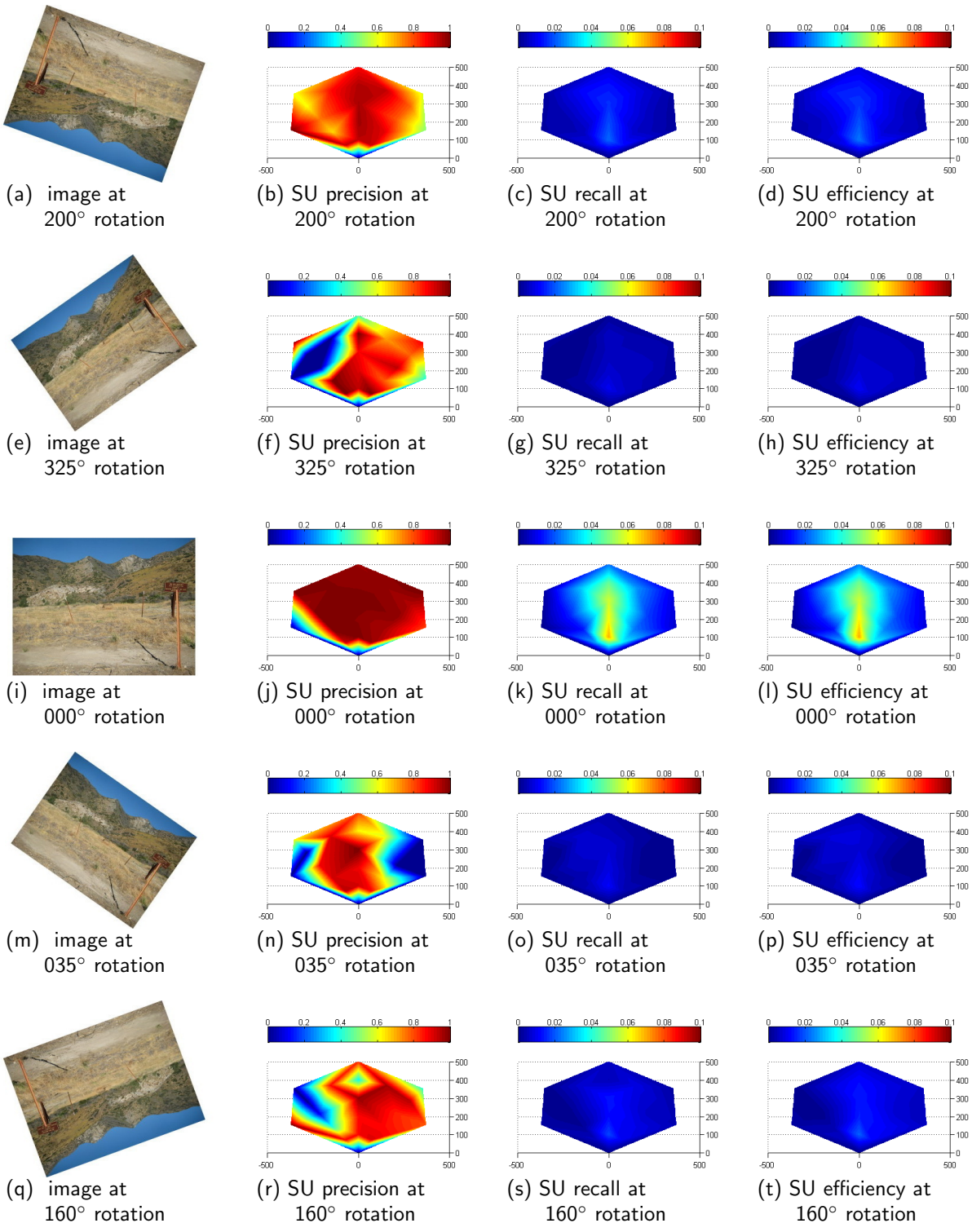
(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

Figure 70. Heat maps for descriptor SC in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



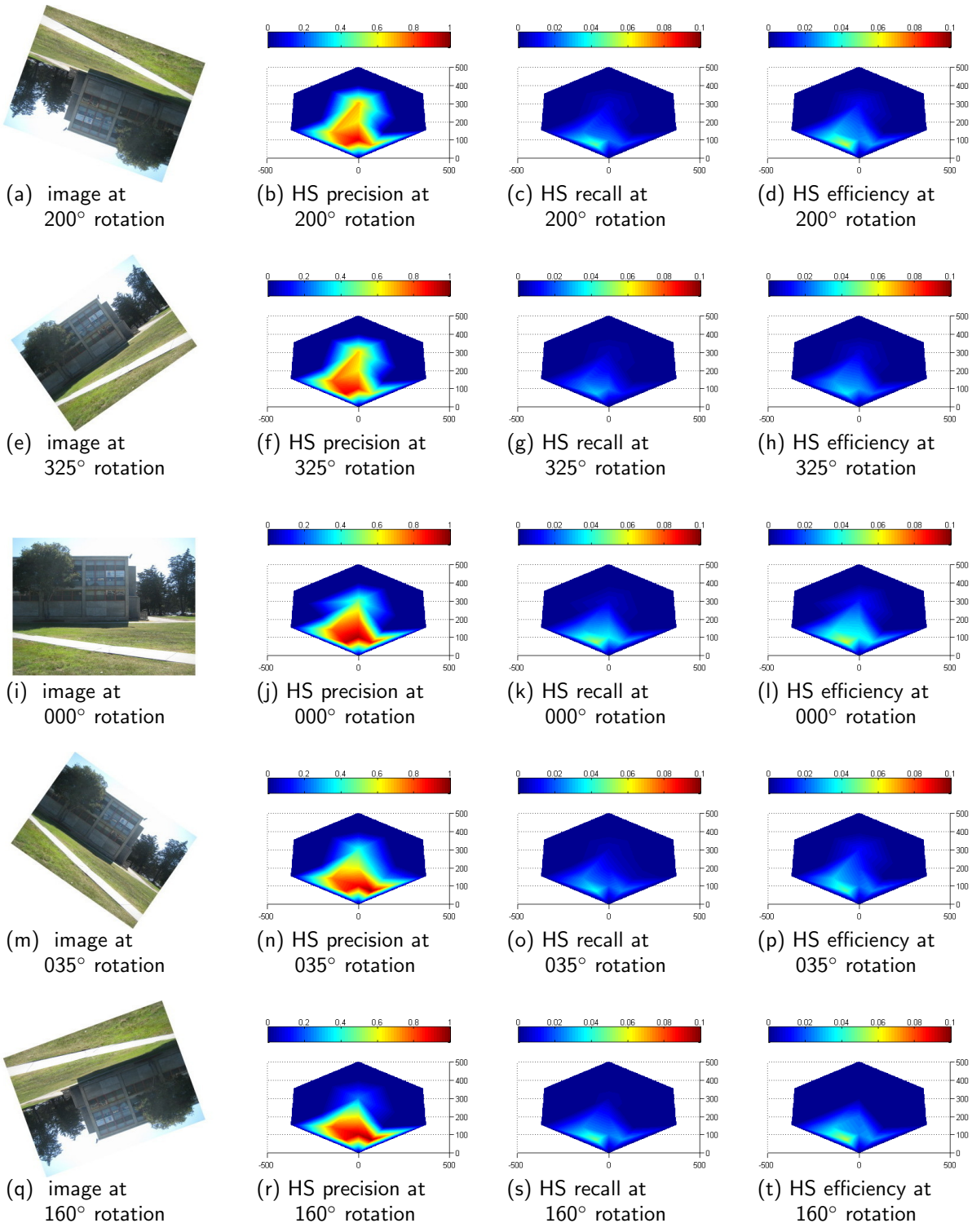
(a) image at 200° rotation
(b) LS precision at 200° rotation
(c) LS recall at 200° rotation
(d) LS efficiency at 200° rotation
(e) image at 325° rotation
(f) LS precision at 325° rotation
(g) LS recall at 325° rotation
(h) LS efficiency at 325° rotation
(i) image at 000° rotation
(j) LS precision at 000° rotation
(k) LS recall at 000° rotation
(l) LS efficiency at 000° rotation
(m) image at 035° rotation
(n) LS precision at 035° rotation
(o) LS recall at 035° rotation
(p) LS efficiency at 035° rotation
(q) image at 160° rotation
(r) LS precision at 160° rotation
(s) LS recall at 160° rotation
(t) LS efficiency at 160° rotation

Figure 71. Heat maps for descriptor LS in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) SU precision at 200° rotation
(c) SU recall at 200° rotation
(d) SU efficiency at 200° rotation
(e) image at 325° rotation
(f) SU precision at 325° rotation
(g) SU recall at 325° rotation
(h) SU efficiency at 325° rotation
(i) image at 000° rotation
(j) SU precision at 000° rotation
(k) SU recall at 000° rotation
(l) SU efficiency at 000° rotation
(m) image at 035° rotation
(n) SU precision at 035° rotation
(o) SU recall at 035° rotation
(p) SU efficiency at 035° rotation
(q) image at 160° rotation
(r) SU precision at 160° rotation
(s) SU recall at 160° rotation
(t) SU efficiency at 160° rotation

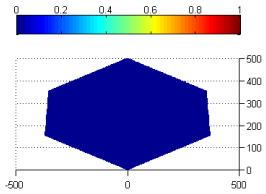
Figure 72. Heat maps for descriptor SU in the OutHay scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



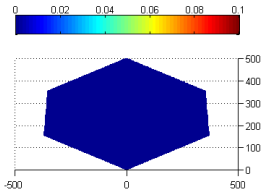
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 73. Heat maps for descriptor HS in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

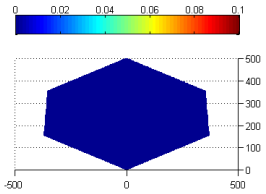
(a) image at 200° rotation



(b) CF precision at 200° rotation

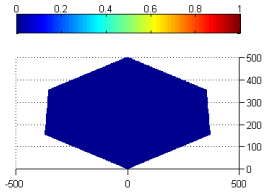


(c) CF recall at 200° rotation

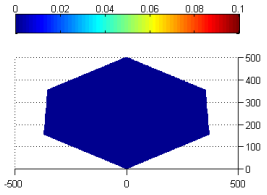


(d) CF efficiency at 200° rotation

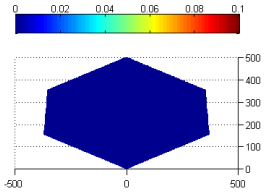
(e) image at 325° rotation



(f) CF precision at 325° rotation

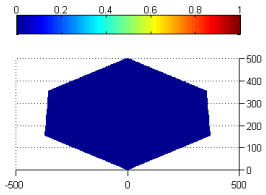


(g) CF recall at 325° rotation

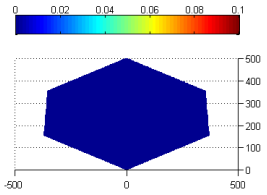


(h) CF efficiency at 325° rotation

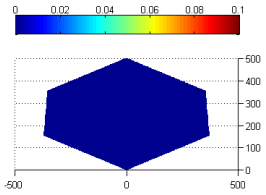
(i) image at 000° rotation



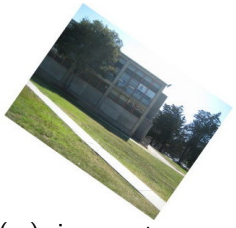
(j) CF precision at 000° rotation



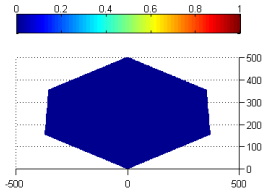
(k) CF recall at 000° rotation



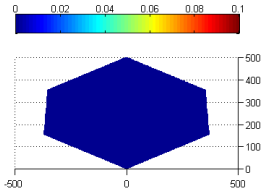
(l) CF efficiency at 000° rotation



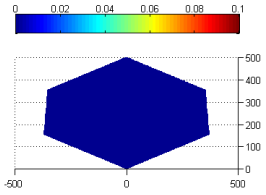
(m) image at 035° rotation



(n) CF precision at 035° rotation

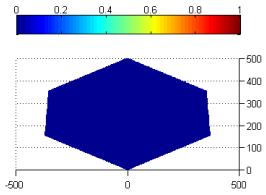


(o) CF recall at 035° rotation

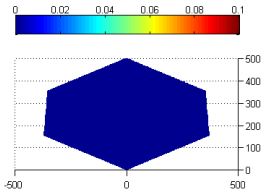


(p) CF efficiency at 035° rotation

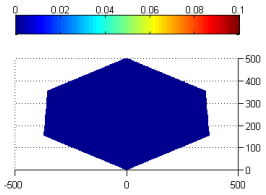
(q) image at 160° rotation



(r) CF precision at 160° rotation



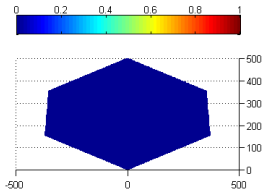
(s) CF recall at 160° rotation



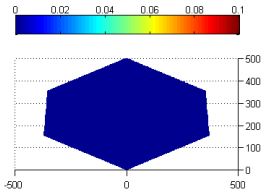
(t) CF efficiency at 160° rotation

Figure 74. Heat maps for descriptor CF in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

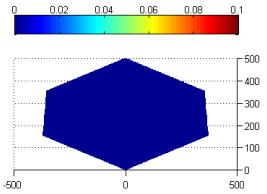
(a) image at 200° rotation



(b) GH precision at 200° rotation

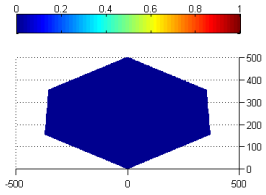


(c) GH recall at 200° rotation

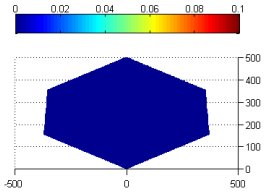


(d) GH efficiency at 200° rotation

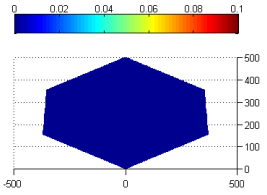
(e) image at 325° rotation



(f) GH precision at 325° rotation

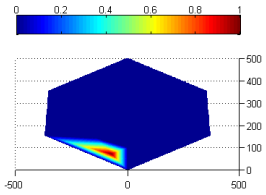


(g) GH recall at 325° rotation

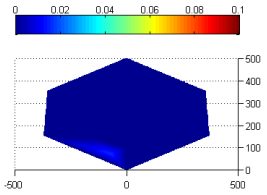


(h) GH efficiency at 325° rotation

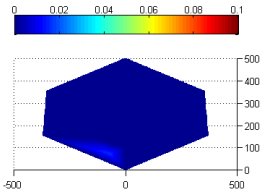
(i) image at 000° rotation



(j) GH precision at 000° rotation

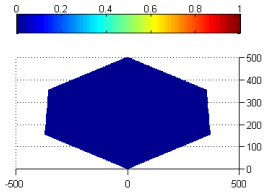


(k) GH recall at 000° rotation

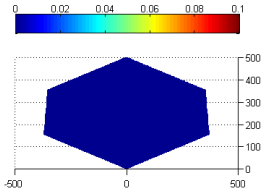


(l) GH efficiency at 000° rotation

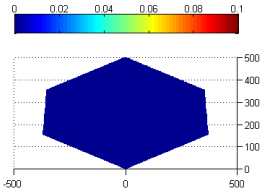
(m) image at 035° rotation



(n) GH precision at 035° rotation

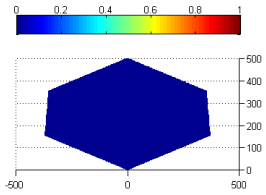


(o) GH recall at 035° rotation

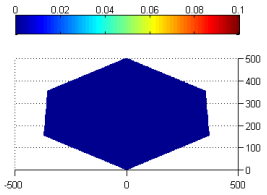


(p) GH efficiency at 035° rotation

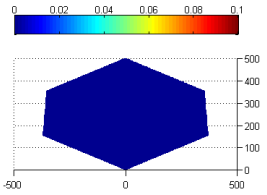
(q) image at 160° rotation



(r) GH precision at 160° rotation

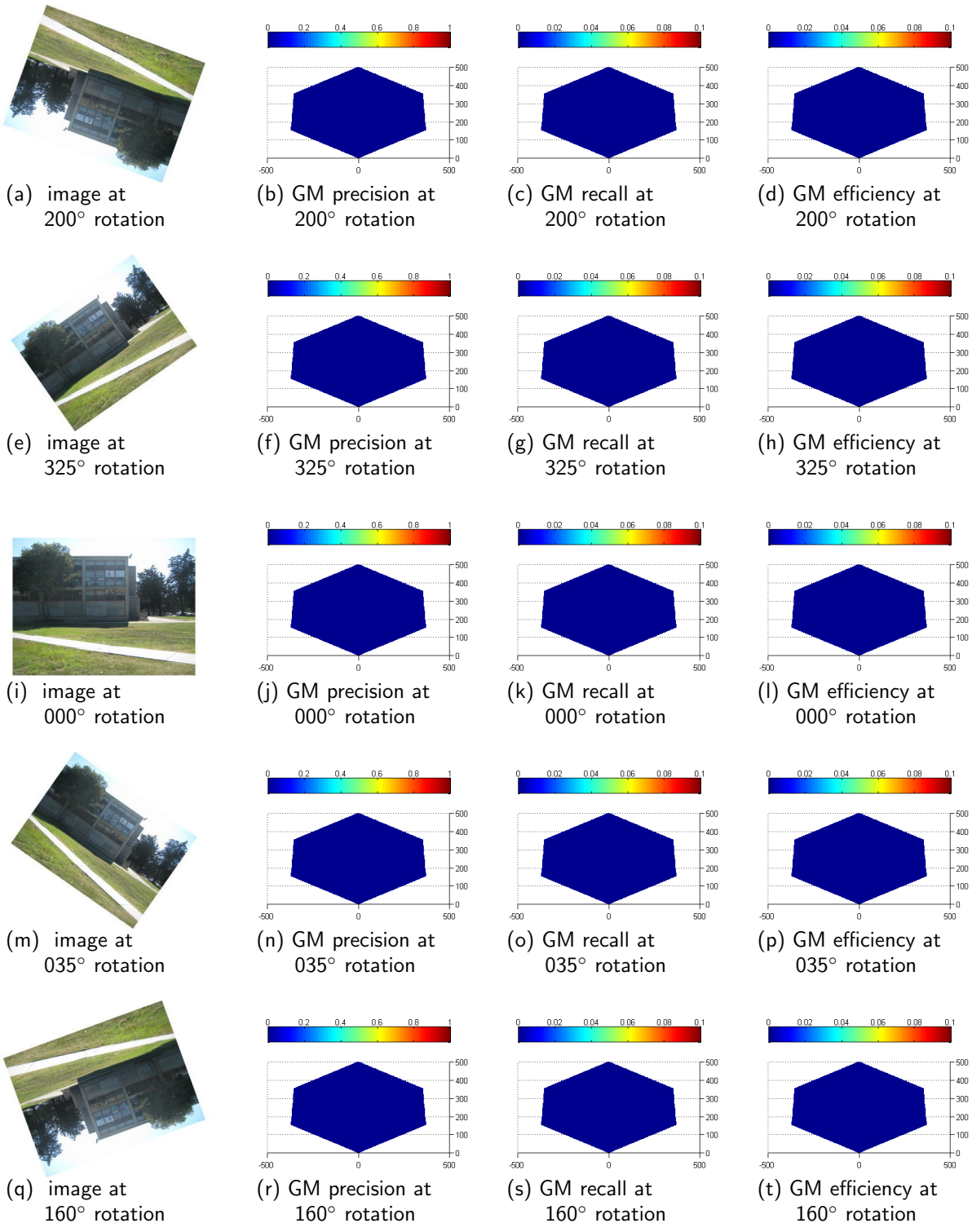


(s) GH recall at 160° rotation



(t) GH efficiency at 160° rotation

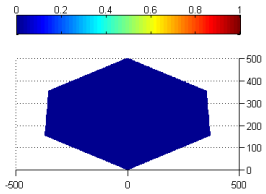
Figure 75. Heat maps for descriptor GH in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



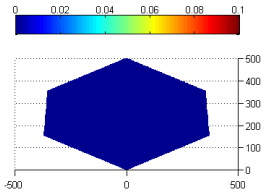
(a) image at 200° rotation
(b) GM precision at 200° rotation
(c) GM recall at 200° rotation
(d) GM efficiency at 200° rotation
(e) image at 325° rotation
(f) GM precision at 325° rotation
(g) GM recall at 325° rotation
(h) GM efficiency at 325° rotation
(i) image at 000° rotation
(j) GM precision at 000° rotation
(k) GM recall at 000° rotation
(l) GM efficiency at 000° rotation
(m) image at 035° rotation
(n) GM precision at 035° rotation
(o) GM recall at 035° rotation
(p) GM efficiency at 035° rotation
(q) image at 160° rotation
(r) GM precision at 160° rotation
(s) GM recall at 160° rotation
(t) GM efficiency at 160° rotation

Figure 76. Heat maps for descriptor GM in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

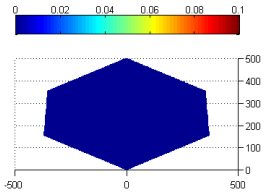
(a) image at 200° rotation



(b) DI precision at 200° rotation

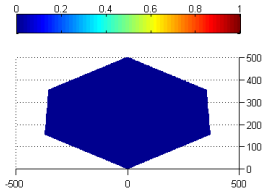


(c) DI recall at 200° rotation

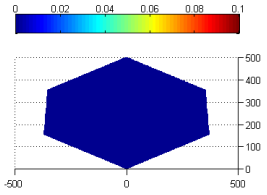


(d) DI efficiency at 200° rotation

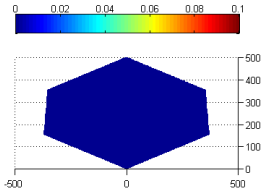
(e) image at 325° rotation



(f) DI precision at 325° rotation

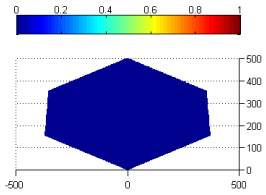


(g) DI recall at 325° rotation

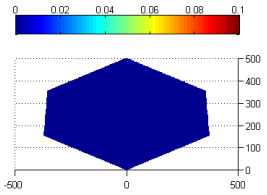


(h) DI efficiency at 325° rotation

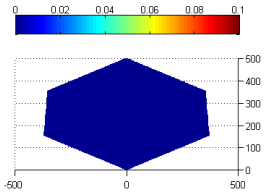
(i) image at 000° rotation



(j) DI precision at 000° rotation

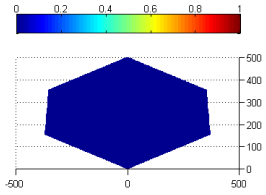


(k) DI recall at 000° rotation

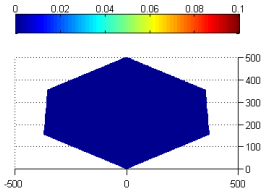


(l) DI efficiency at 000° rotation

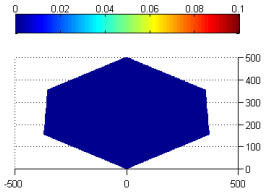
(m) image at 035° rotation



(n) DI precision at 035° rotation

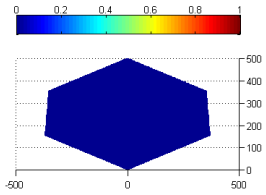


(o) DI recall at 035° rotation

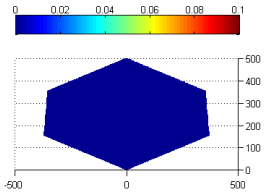


(p) DI efficiency at 035° rotation

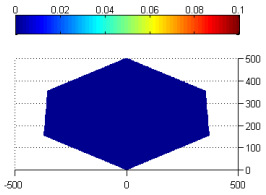
(q) image at 160° rotation



(r) DI precision at 160° rotation



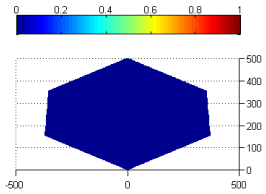
(s) DI recall at 160° rotation



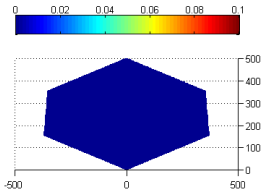
(t) DI efficiency at 160° rotation

Figure 77. Heat maps for descriptor DI in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

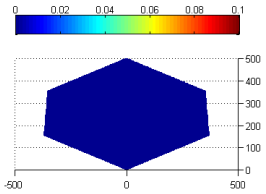
(a) image at 200° rotation



(b) CC precision at 200° rotation

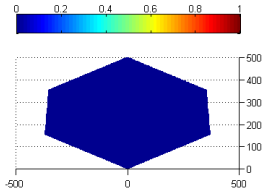


(c) CC recall at 200° rotation

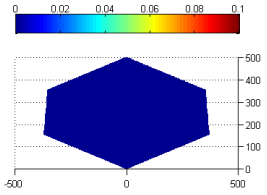


(d) CC efficiency at 200° rotation

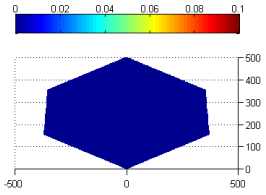
(e) image at 325° rotation



(f) CC precision at 325° rotation

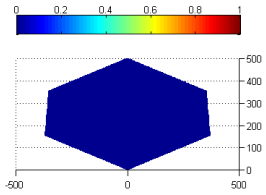


(g) CC recall at 325° rotation

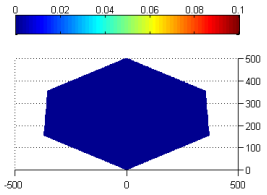


(h) CC efficiency at 325° rotation

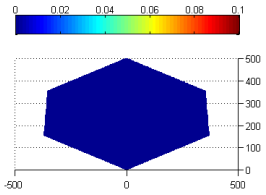
(i) image at 000° rotation



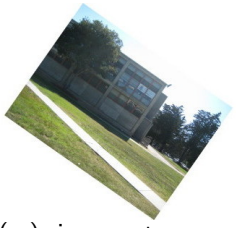
(j) CC precision at 000° rotation



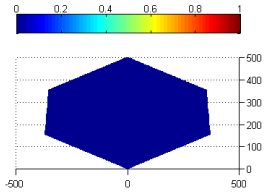
(k) CC recall at 000° rotation



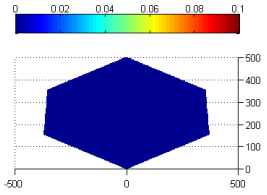
(l) CC efficiency at 000° rotation



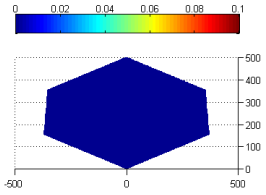
(m) image at 035° rotation



(n) CC precision at 035° rotation

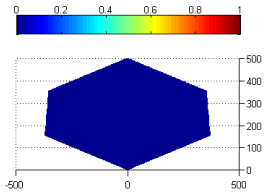


(o) CC recall at 035° rotation

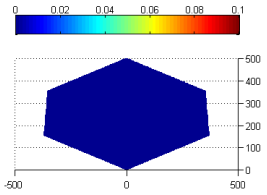


(p) CC efficiency at 035° rotation

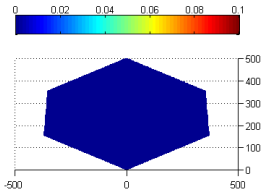
(q) image at 160° rotation



(r) CC precision at 160° rotation

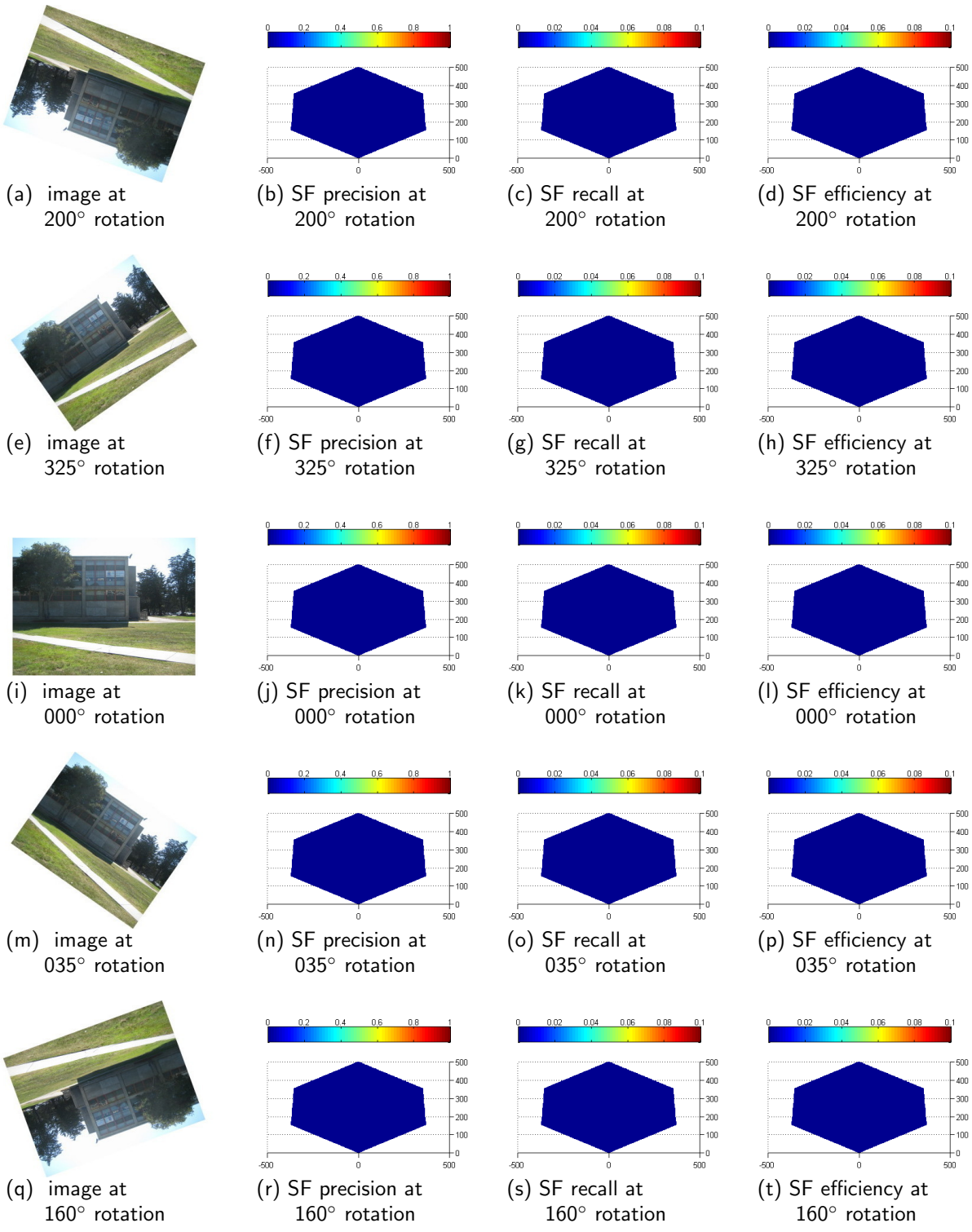


(s) CC recall at 160° rotation



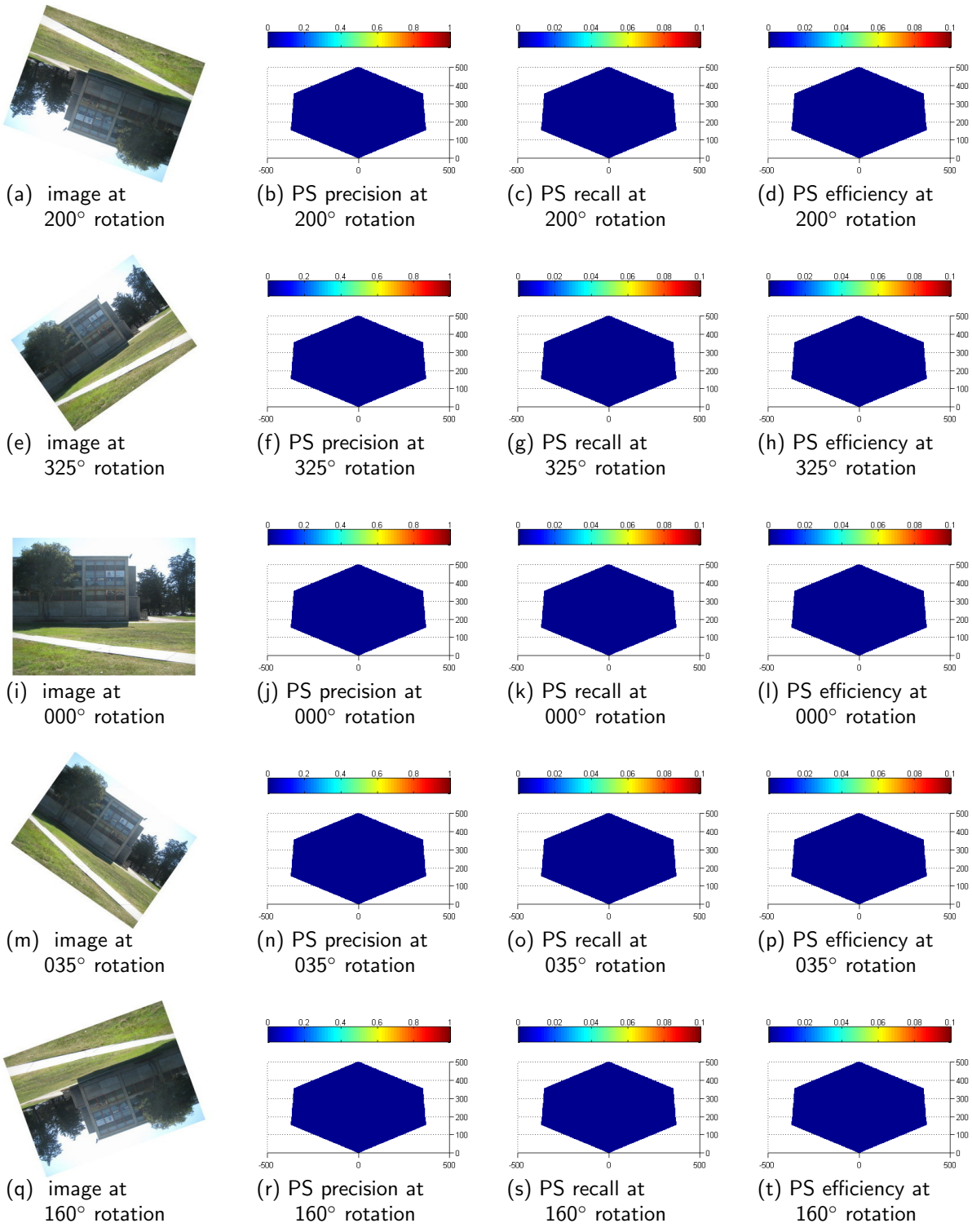
(t) CC efficiency at 160° rotation

Figure 78. Heat maps for descriptor CC in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



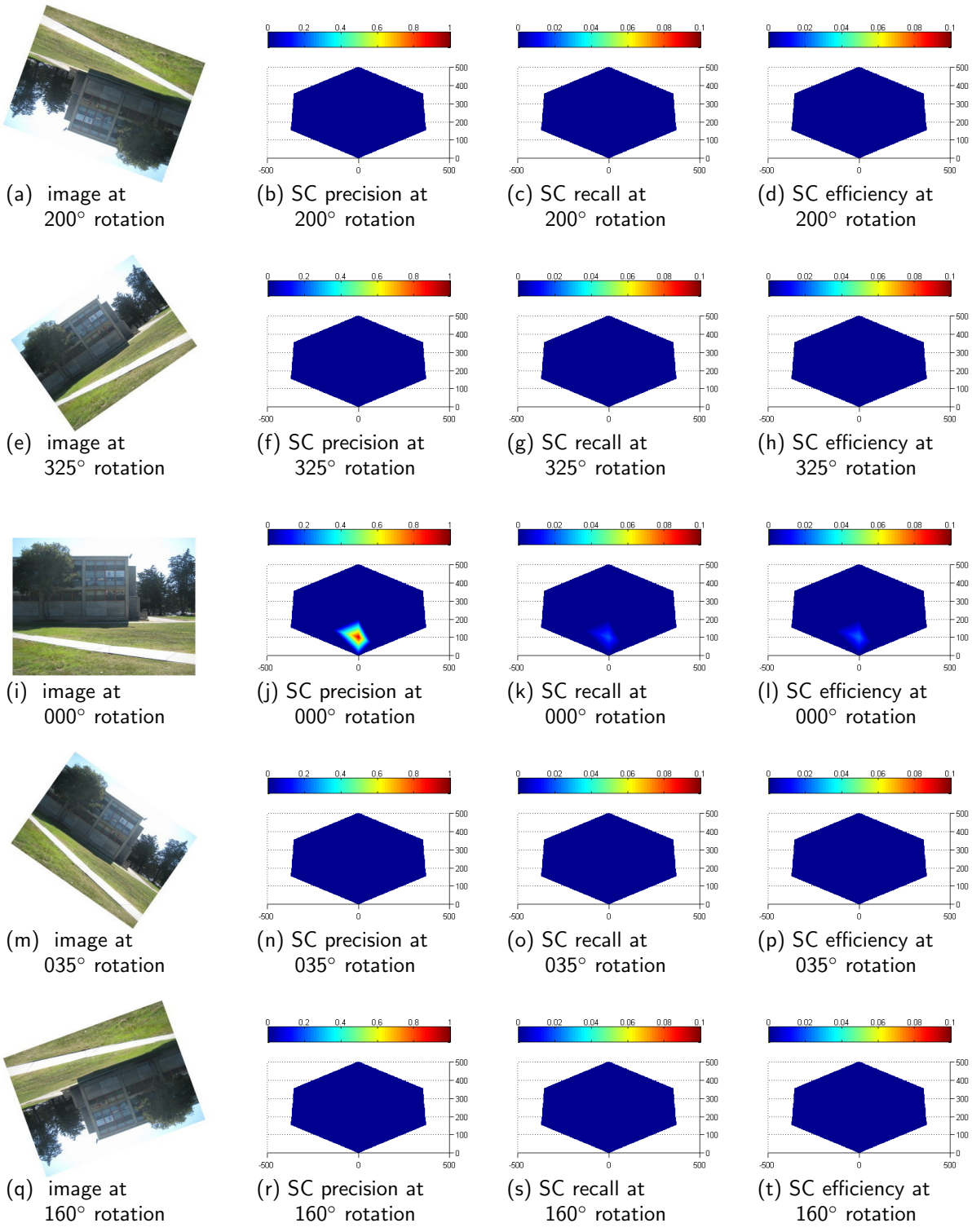
(a) image at 200° rotation
(b) SF precision at 200° rotation
(c) SF recall at 200° rotation
(d) SF efficiency at 200° rotation
(e) image at 325° rotation
(f) SF precision at 325° rotation
(g) SF recall at 325° rotation
(h) SF efficiency at 325° rotation
(i) image at 000° rotation
(j) SF precision at 000° rotation
(k) SF recall at 000° rotation
(l) SF efficiency at 000° rotation
(m) image at 035° rotation
(n) SF precision at 035° rotation
(o) SF recall at 035° rotation
(p) SF efficiency at 035° rotation
(q) image at 160° rotation
(r) SF precision at 160° rotation
(s) SF recall at 160° rotation
(t) SF efficiency at 160° rotation

Figure 79. Heat maps for descriptor SF in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



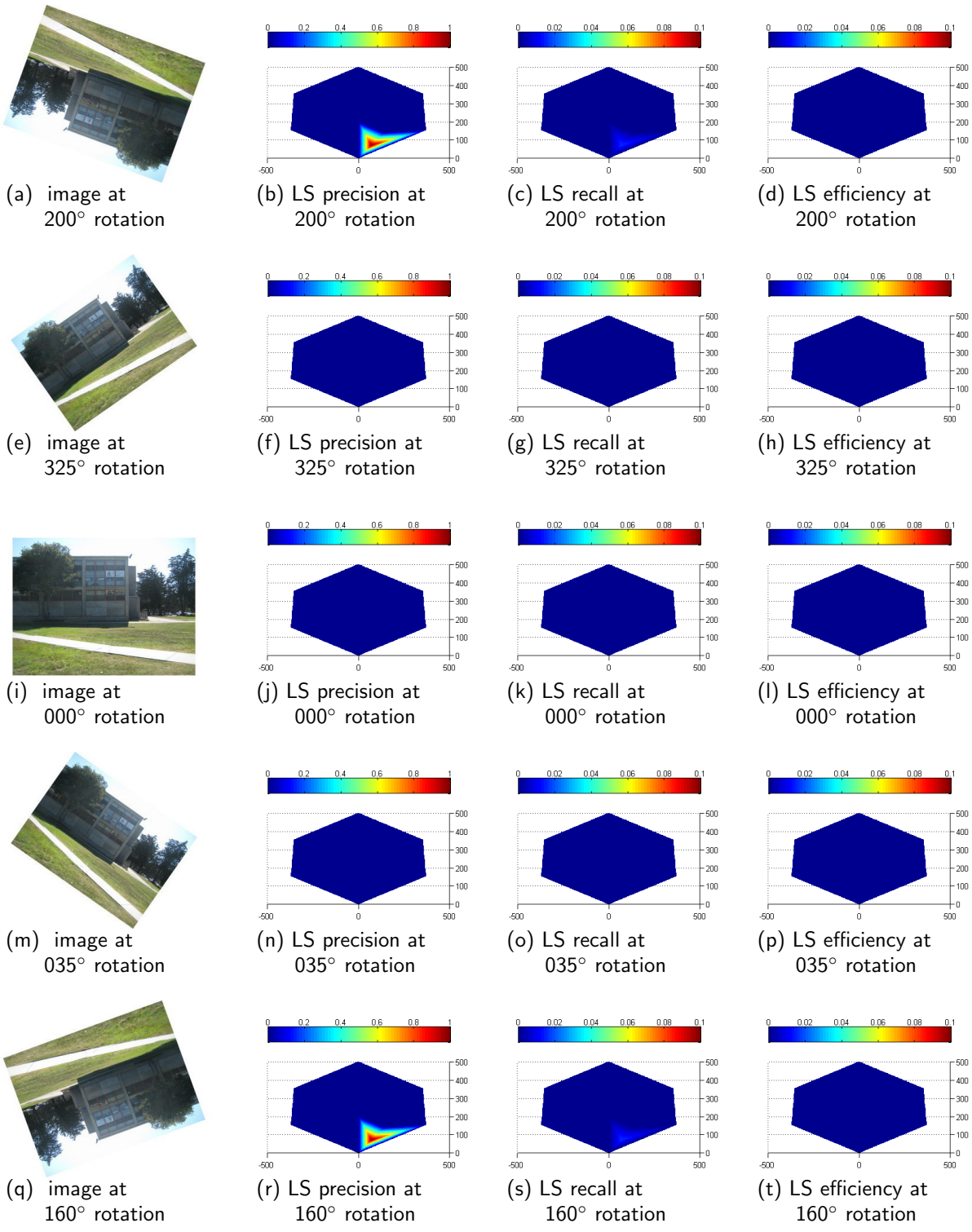
(a) image at 200° rotation
(b) PS precision at 200° rotation
(c) PS recall at 200° rotation
(d) PS efficiency at 200° rotation
(e) image at 325° rotation
(f) PS precision at 325° rotation
(g) PS recall at 325° rotation
(h) PS efficiency at 325° rotation
(i) image at 000° rotation
(j) PS precision at 000° rotation
(k) PS recall at 000° rotation
(l) PS efficiency at 000° rotation
(m) image at 035° rotation
(n) PS precision at 035° rotation
(o) PS recall at 035° rotation
(p) PS efficiency at 035° rotation
(q) image at 160° rotation
(r) PS precision at 160° rotation
(s) PS recall at 160° rotation
(t) PS efficiency at 160° rotation

Figure 80. Heat maps for descriptor PS in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

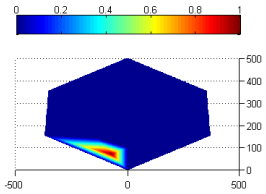
Figure 81. Heat maps for descriptor SC in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



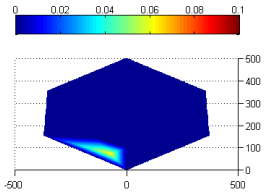
(a) image at 200° rotation
(b) LS precision at 200° rotation
(c) LS recall at 200° rotation
(d) LS efficiency at 200° rotation
(e) image at 325° rotation
(f) LS precision at 325° rotation
(g) LS recall at 325° rotation
(h) LS efficiency at 325° rotation
(i) image at 000° rotation
(j) LS precision at 000° rotation
(k) LS recall at 000° rotation
(l) LS efficiency at 000° rotation
(m) image at 035° rotation
(n) LS precision at 035° rotation
(o) LS recall at 035° rotation
(p) LS efficiency at 035° rotation
(q) image at 160° rotation
(r) LS precision at 160° rotation
(s) LS recall at 160° rotation
(t) LS efficiency at 160° rotation

Figure 82. Heat maps for descriptor LS in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

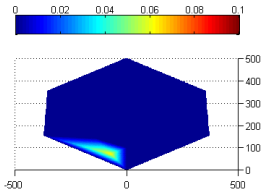
(a) image at 200° rotation



(b) SU precision at 200° rotation

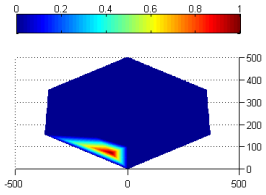


(c) SU recall at 200° rotation

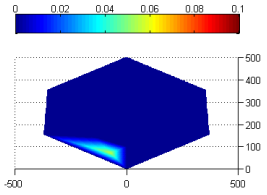


(d) SU efficiency at 200° rotation

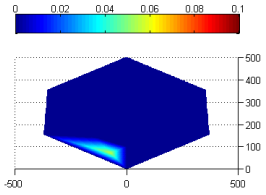
(e) image at 325° rotation



(f) SU precision at 325° rotation

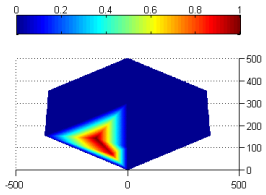


(g) SU recall at 325° rotation

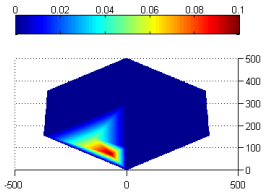


(h) SU efficiency at 325° rotation

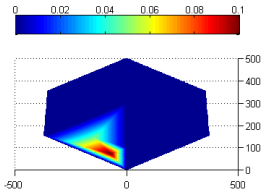
(i) image at 000° rotation



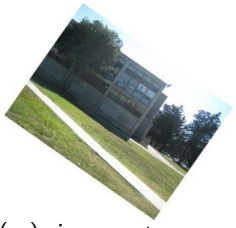
(j) SU precision at 000° rotation



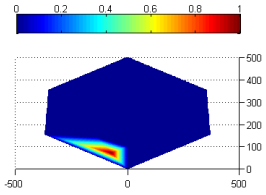
(k) SU recall at 000° rotation



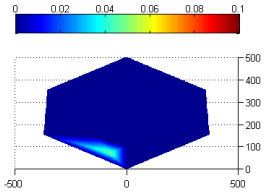
(l) SU efficiency at 000° rotation



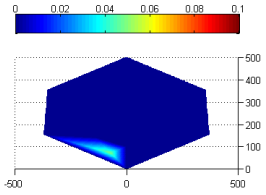
(m) image at 035° rotation



(n) SU precision at 035° rotation

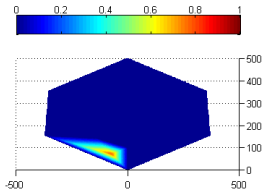


(o) SU recall at 035° rotation

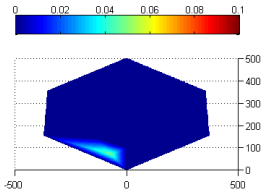


(p) SU efficiency at 035° rotation

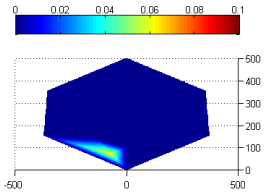
(q) image at 160° rotation



(r) SU precision at 160° rotation

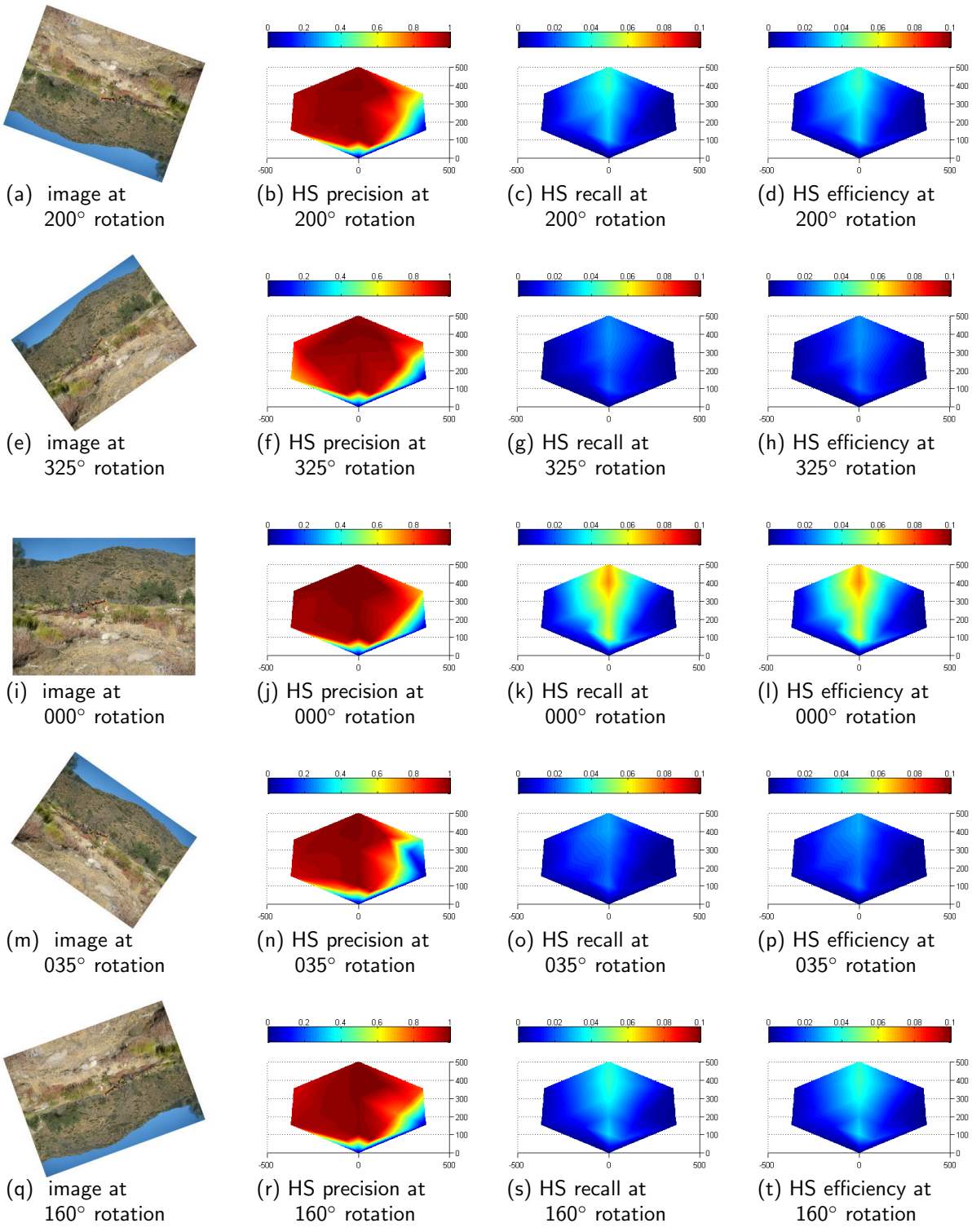


(s) SU recall at 160° rotation



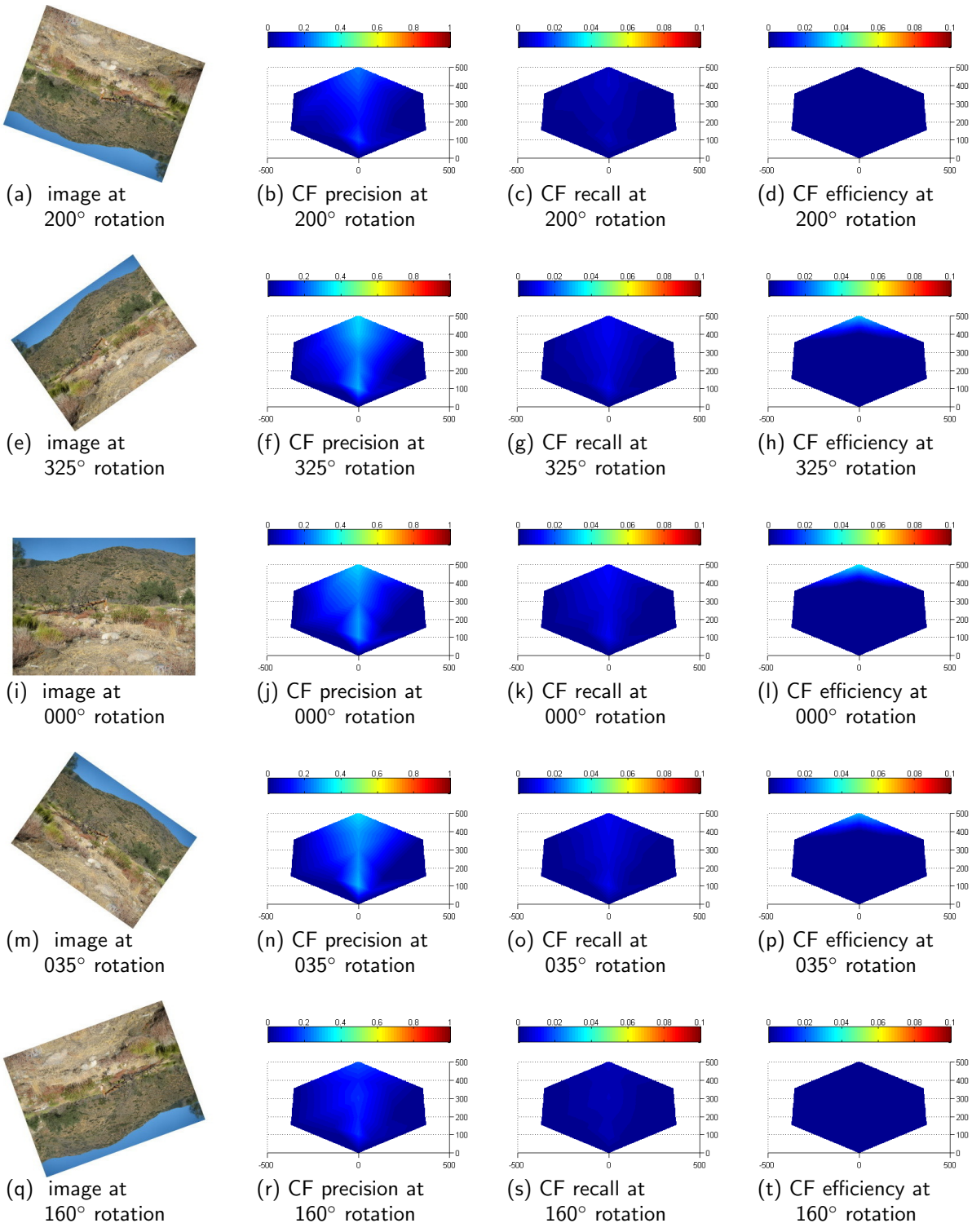
(t) SU efficiency at 160° rotation

Figure 83. Heat maps for descriptor SU in the OutHaligan scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



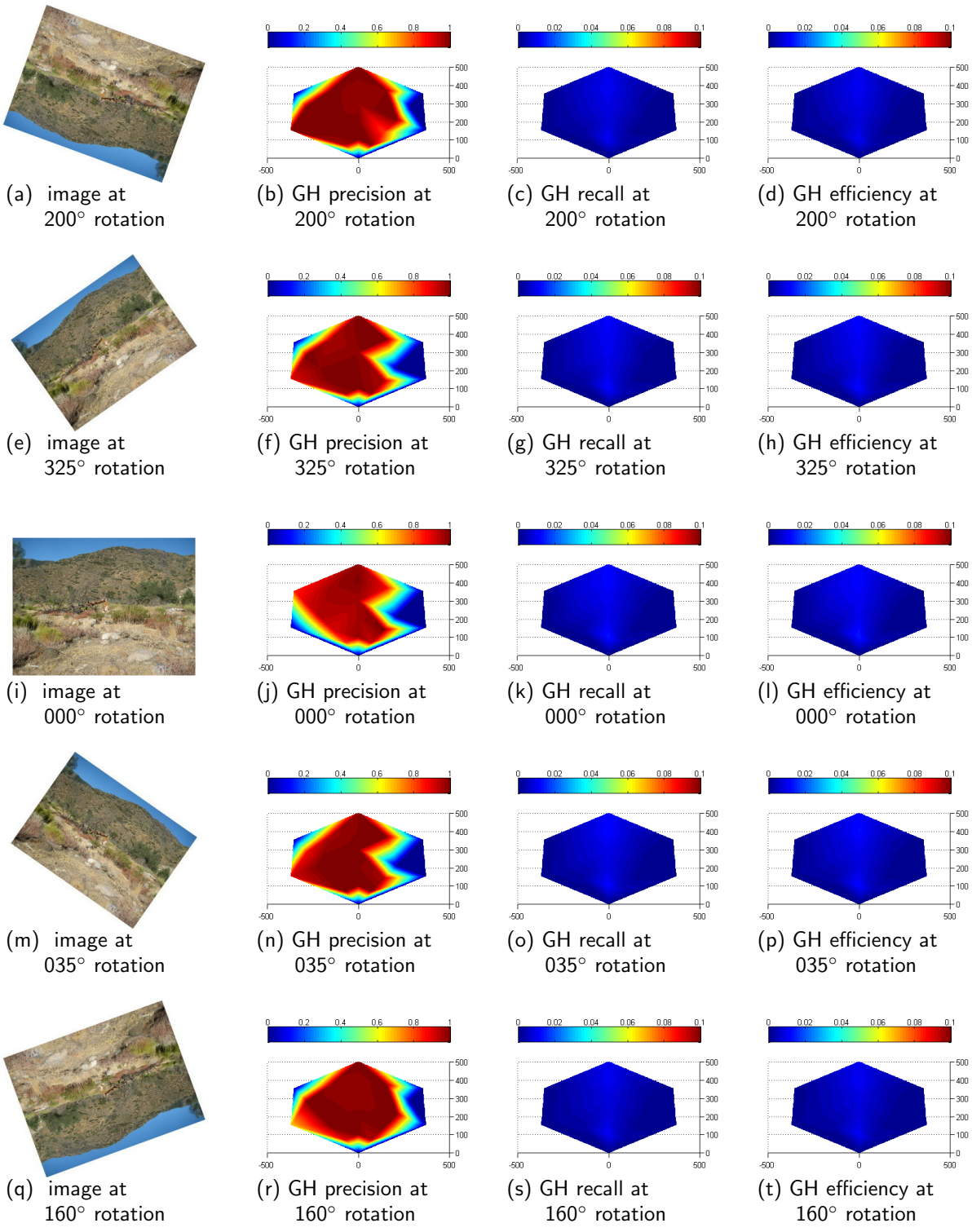
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 84. Heat maps for descriptor HS in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



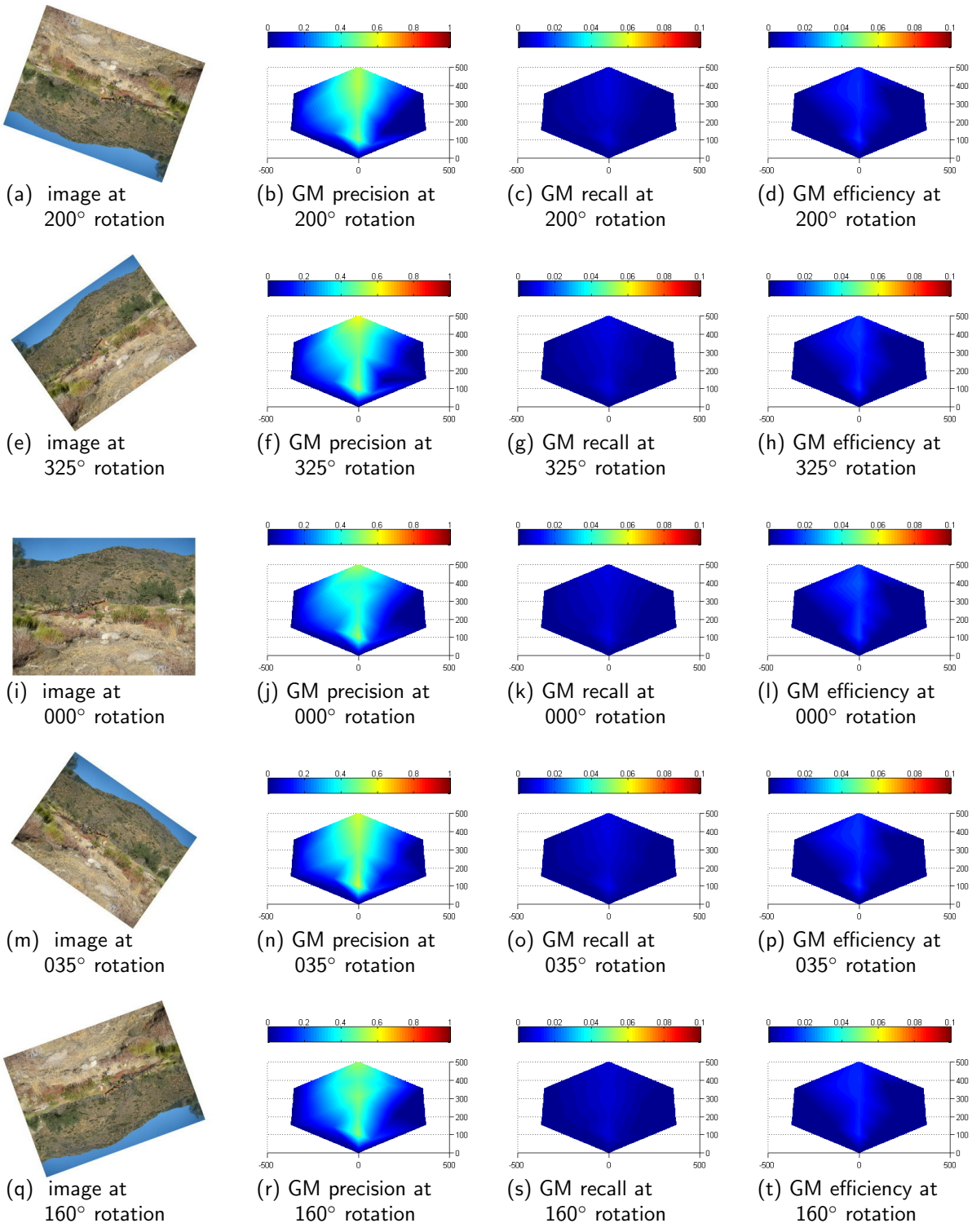
(a) image at 200° rotation
(b) CF precision at 200° rotation
(c) CF recall at 200° rotation
(d) CF efficiency at 200° rotation
(e) image at 325° rotation
(f) CF precision at 325° rotation
(g) CF recall at 325° rotation
(h) CF efficiency at 325° rotation
(i) image at 000° rotation
(j) CF precision at 000° rotation
(k) CF recall at 000° rotation
(l) CF efficiency at 000° rotation
(m) image at 035° rotation
(n) CF precision at 035° rotation
(o) CF recall at 035° rotation
(p) CF efficiency at 035° rotation
(q) image at 160° rotation
(r) CF precision at 160° rotation
(s) CF recall at 160° rotation
(t) CF efficiency at 160° rotation

Figure 85. Heat maps for descriptor CF in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



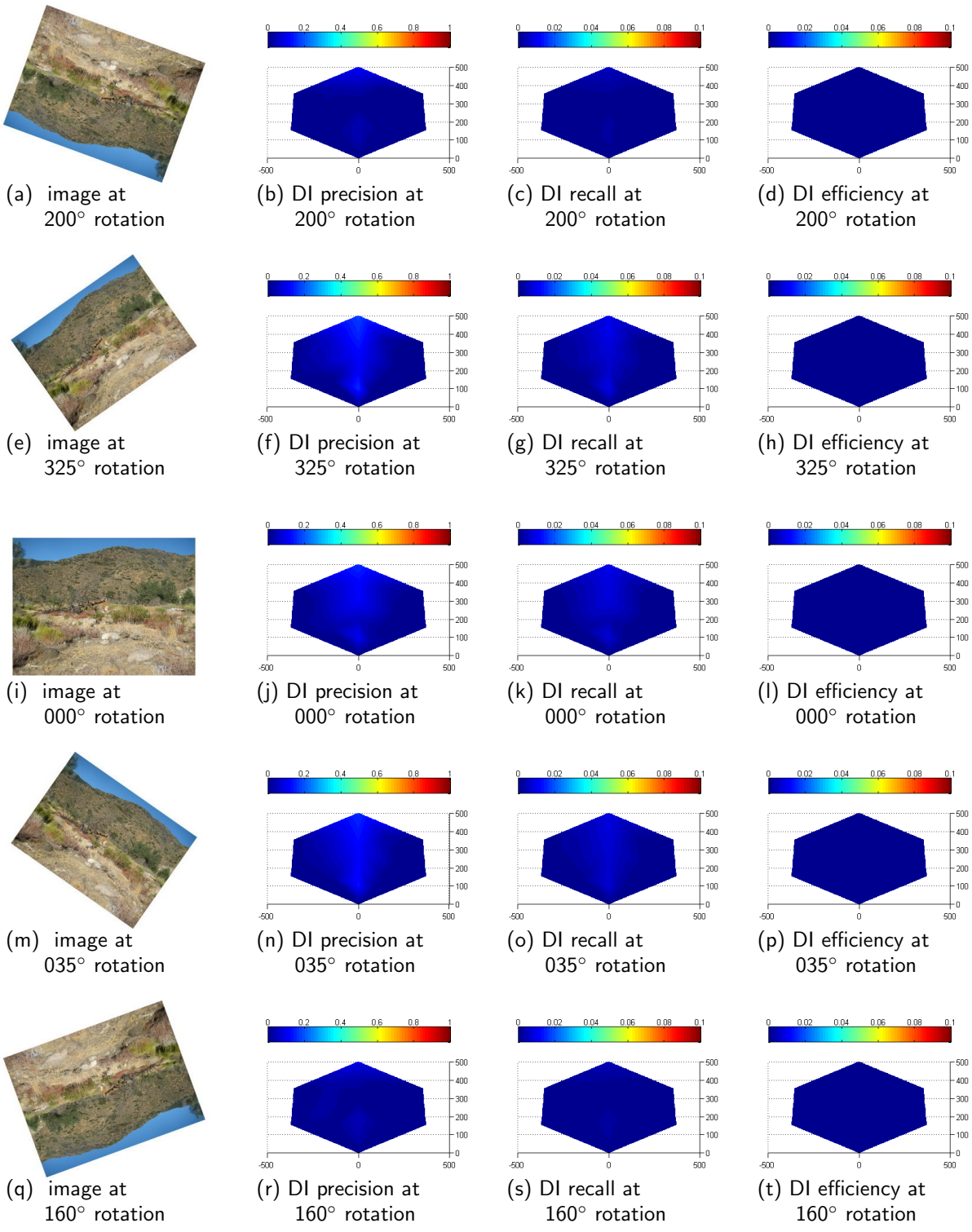
(a) image at 200° rotation
(b) GH precision at 200° rotation
(c) GH recall at 200° rotation
(d) GH efficiency at 200° rotation
(e) image at 325° rotation
(f) GH precision at 325° rotation
(g) GH recall at 325° rotation
(h) GH efficiency at 325° rotation
(i) image at 000° rotation
(j) GH precision at 000° rotation
(k) GH recall at 000° rotation
(l) GH efficiency at 000° rotation
(m) image at 035° rotation
(n) GH precision at 035° rotation
(o) GH recall at 035° rotation
(p) GH efficiency at 035° rotation
(q) image at 160° rotation
(r) GH precision at 160° rotation
(s) GH recall at 160° rotation
(t) GH efficiency at 160° rotation

Figure 86. Heat maps for descriptor GH in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



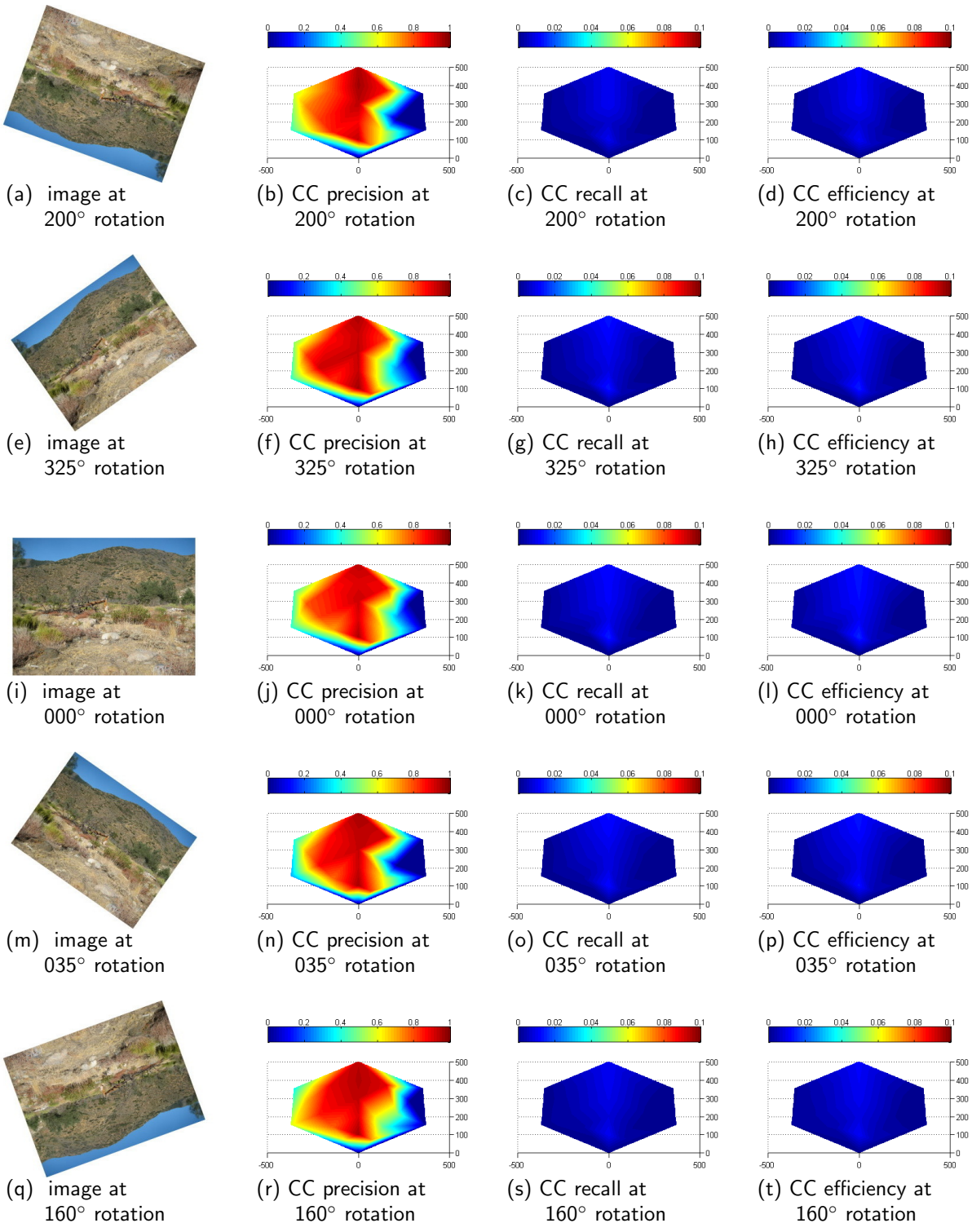
(a) image at 200° rotation
(b) GM precision at 200° rotation
(c) GM recall at 200° rotation
(d) GM efficiency at 200° rotation
(e) image at 325° rotation
(f) GM precision at 325° rotation
(g) GM recall at 325° rotation
(h) GM efficiency at 325° rotation
(i) image at 000° rotation
(j) GM precision at 000° rotation
(k) GM recall at 000° rotation
(l) GM efficiency at 000° rotation
(m) image at 035° rotation
(n) GM precision at 035° rotation
(o) GM recall at 035° rotation
(p) GM efficiency at 035° rotation
(q) image at 160° rotation
(r) GM precision at 160° rotation
(s) GM recall at 160° rotation
(t) GM efficiency at 160° rotation

Figure 87. Heat maps for descriptor GM in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



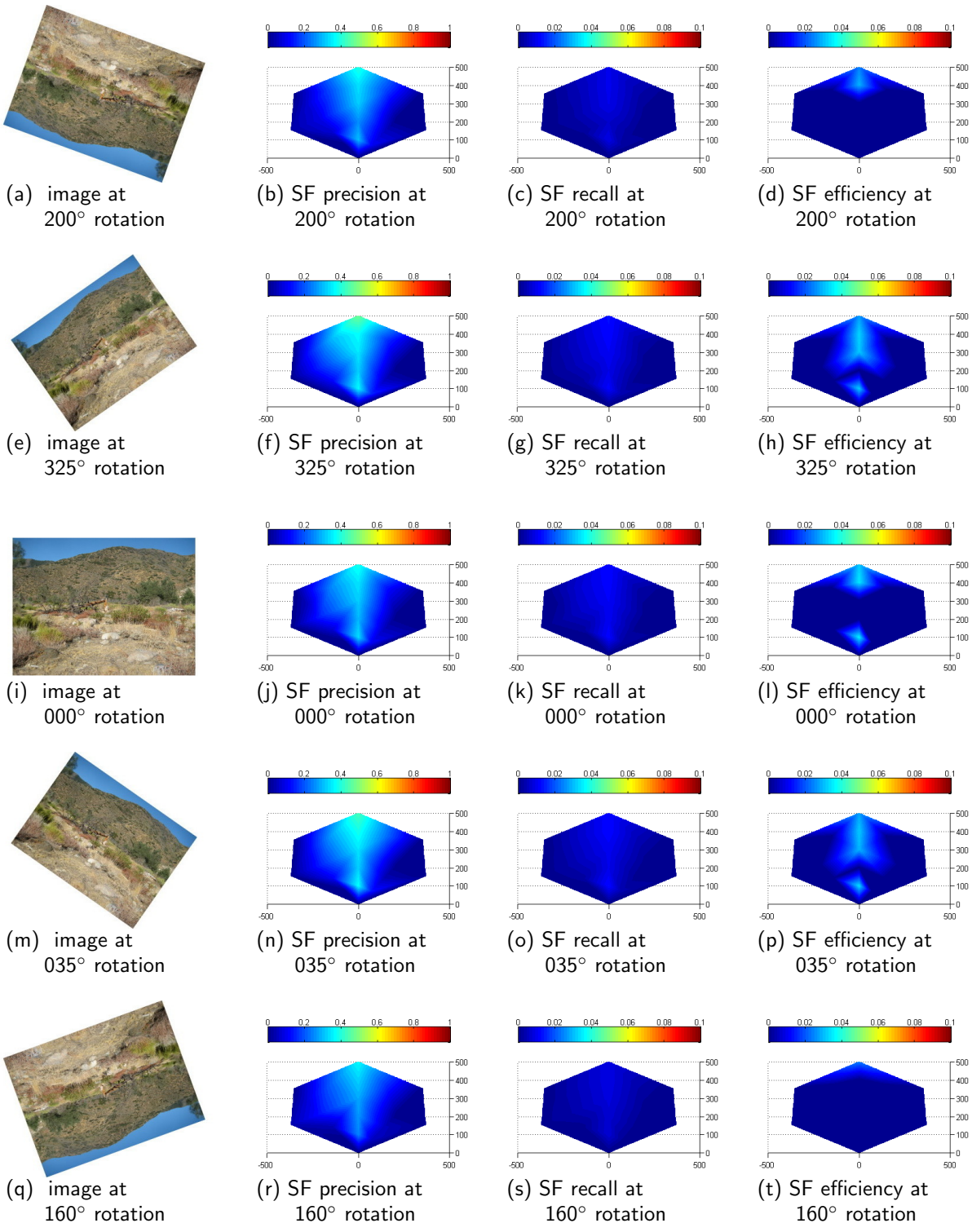
(a) image at 200° rotation
(b) DI precision at 200° rotation
(c) DI recall at 200° rotation
(d) DI efficiency at 200° rotation
(e) image at 325° rotation
(f) DI precision at 325° rotation
(g) DI recall at 325° rotation
(h) DI efficiency at 325° rotation
(i) image at 000° rotation
(j) DI precision at 000° rotation
(k) DI recall at 000° rotation
(l) DI efficiency at 000° rotation
(m) image at 035° rotation
(n) DI precision at 035° rotation
(o) DI recall at 035° rotation
(p) DI efficiency at 035° rotation
(q) image at 160° rotation
(r) DI precision at 160° rotation
(s) DI recall at 160° rotation
(t) DI efficiency at 160° rotation

Figure 88. Heat maps for descriptor DI in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) CC precision at 200° rotation
(c) CC recall at 200° rotation
(d) CC efficiency at 200° rotation
(e) image at 325° rotation
(f) CC precision at 325° rotation
(g) CC recall at 325° rotation
(h) CC efficiency at 325° rotation
(i) image at 000° rotation
(j) CC precision at 000° rotation
(k) CC recall at 000° rotation
(l) CC efficiency at 000° rotation
(m) image at 035° rotation
(n) CC precision at 035° rotation
(o) CC recall at 035° rotation
(p) CC efficiency at 035° rotation
(q) image at 160° rotation
(r) CC precision at 160° rotation
(s) CC recall at 160° rotation
(t) CC efficiency at 160° rotation

Figure 89. Heat maps for descriptor CC in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) SF precision at 200° rotation
(c) SF recall at 200° rotation
(d) SF efficiency at 200° rotation
(e) image at 325° rotation
(f) SF precision at 325° rotation
(g) SF recall at 325° rotation
(h) SF efficiency at 325° rotation
(i) image at 000° rotation
(j) SF precision at 000° rotation
(k) SF recall at 000° rotation
(l) SF efficiency at 000° rotation
(m) image at 035° rotation
(n) SF precision at 035° rotation
(o) SF recall at 035° rotation
(p) SF efficiency at 035° rotation
(q) image at 160° rotation
(r) SF precision at 160° rotation
(s) SF recall at 160° rotation
(t) SF efficiency at 160° rotation

Figure 90. Heat maps for descriptor SF in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

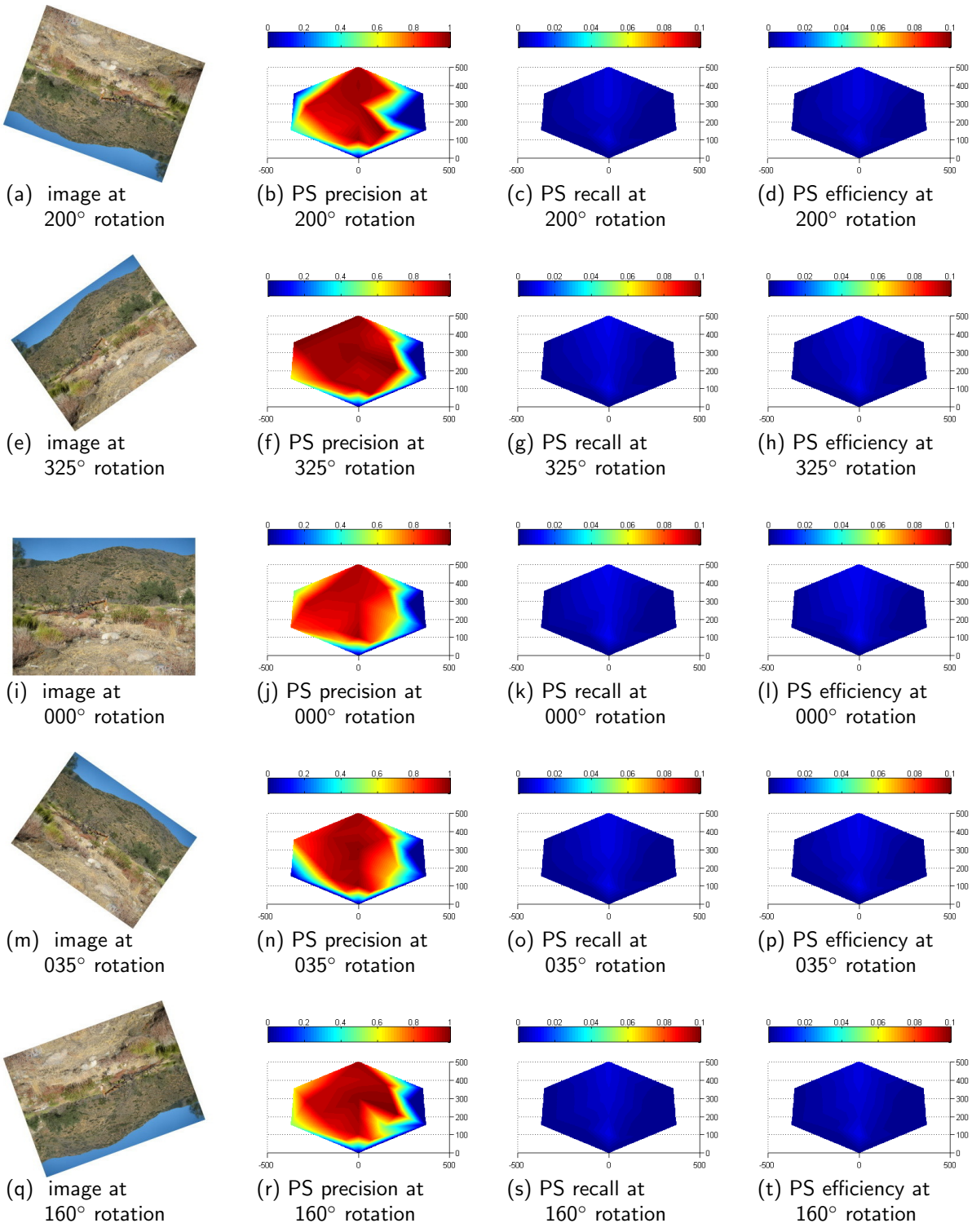
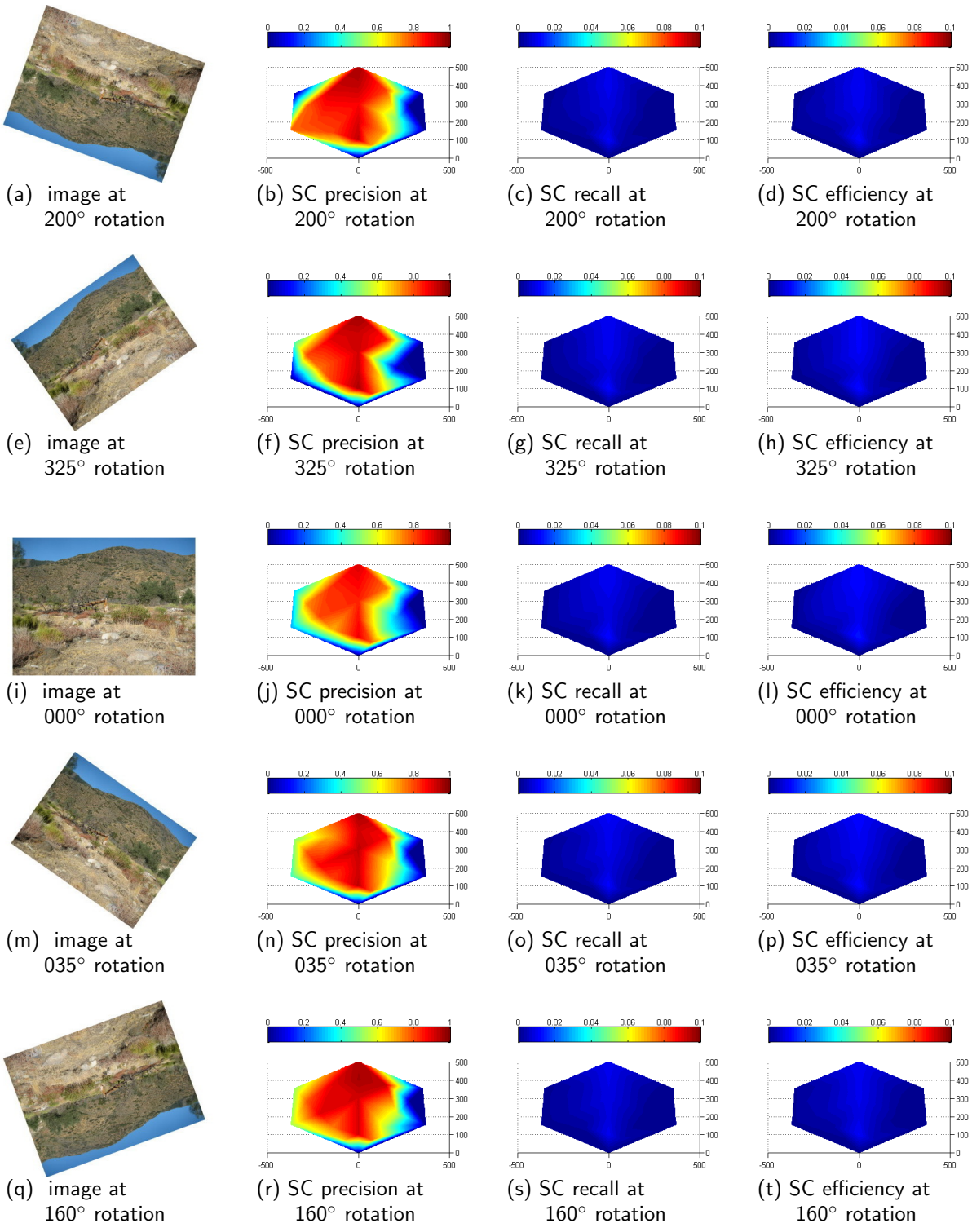
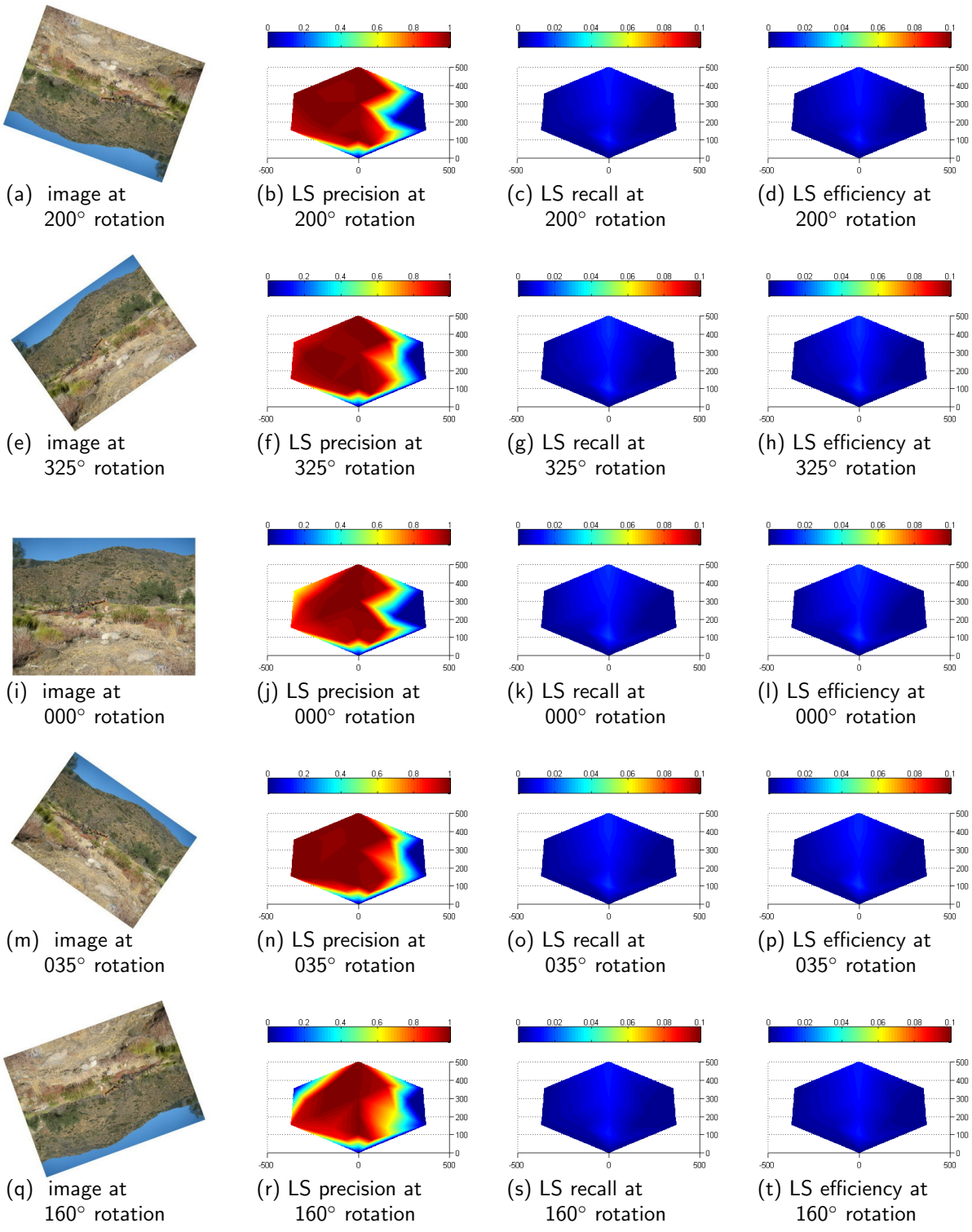


Figure 91. Heat maps for descriptor PS in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

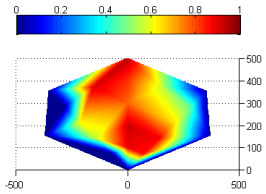
Figure 92. Heat maps for descriptor SC in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



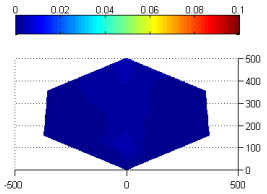
(a) image at 200° rotation
(b) LS precision at 200° rotation
(c) LS recall at 200° rotation
(d) LS efficiency at 200° rotation
(e) image at 325° rotation
(f) LS precision at 325° rotation
(g) LS recall at 325° rotation
(h) LS efficiency at 325° rotation
(i) image at 000° rotation
(j) LS precision at 000° rotation
(k) LS recall at 000° rotation
(l) LS efficiency at 000° rotation
(m) image at 035° rotation
(n) LS precision at 035° rotation
(o) LS recall at 035° rotation
(p) LS efficiency at 035° rotation
(q) image at 160° rotation
(r) LS precision at 160° rotation
(s) LS recall at 160° rotation
(t) LS efficiency at 160° rotation

Figure 93. Heat maps for descriptor LS in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

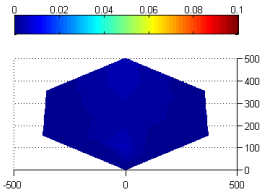
(a) image at 200° rotation



(b) SU precision at 200° rotation

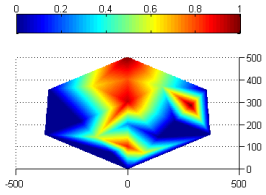


(c) SU recall at 200° rotation

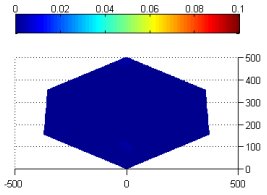


(d) SU efficiency at 200° rotation

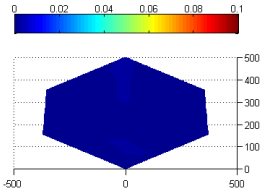
(e) image at 325° rotation



(f) SU precision at 325° rotation

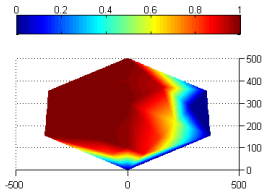


(g) SU recall at 325° rotation

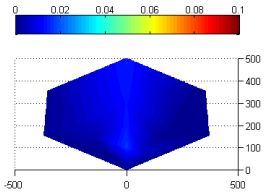


(h) SU efficiency at 325° rotation

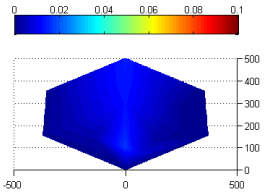
(i) image at 000° rotation



(j) SU precision at 000° rotation

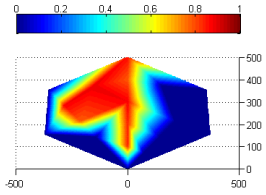


(k) SU recall at 000° rotation

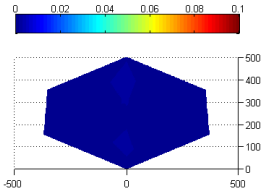


(l) SU efficiency at 000° rotation

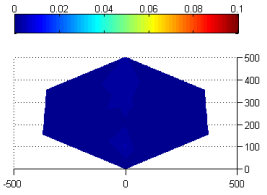
(m) image at 035° rotation



(n) SU precision at 035° rotation

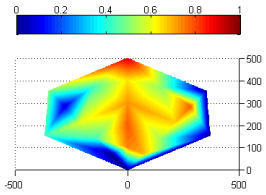


(o) SU recall at 035° rotation

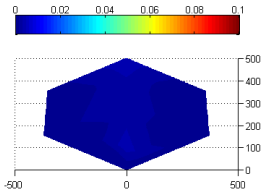


(p) SU efficiency at 035° rotation

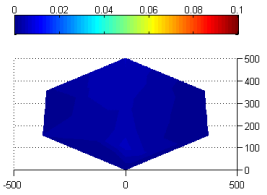
(q) image at 160° rotation



(r) SU precision at 160° rotation

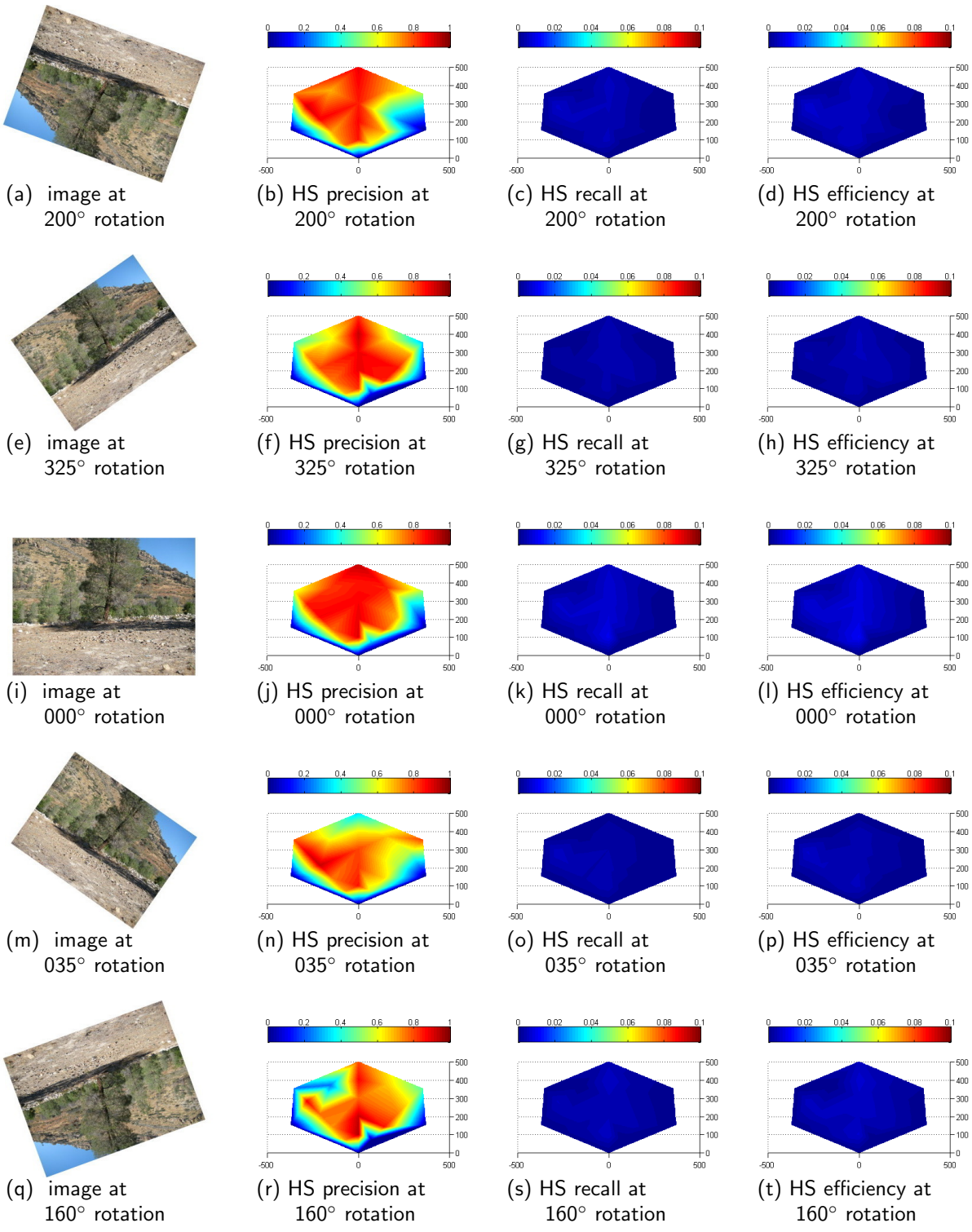


(s) SU recall at 160° rotation



(t) SU efficiency at 160° rotation

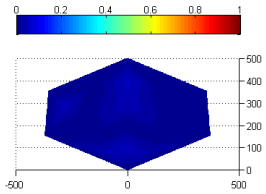
Figure 94. Heat maps for descriptor SU in the OutStump scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



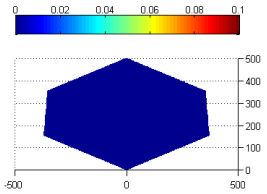
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 95. Heat maps for descriptor HS in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

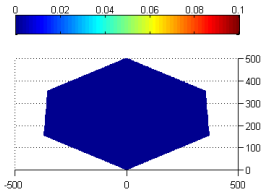
(a) image at 200° rotation



(b) CF precision at 200° rotation

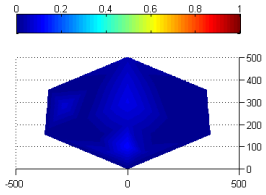


(c) CF recall at 200° rotation

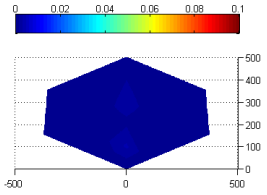


(d) CF efficiency at 200° rotation

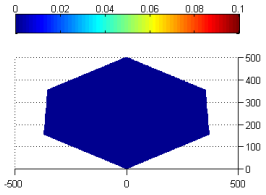
(e) image at 325° rotation



(f) CF precision at 325° rotation

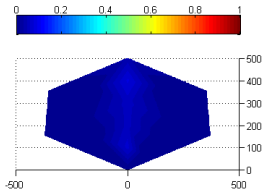


(g) CF recall at 325° rotation

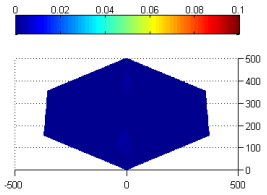


(h) CF efficiency at 325° rotation

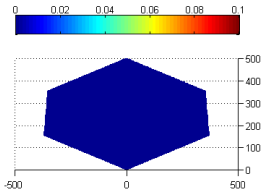
(i) image at 000° rotation



(j) CF precision at 000° rotation

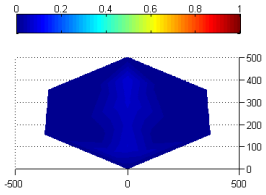


(k) CF recall at 000° rotation

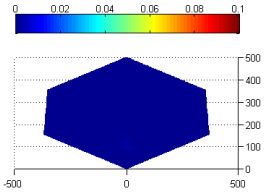


(l) CF efficiency at 000° rotation

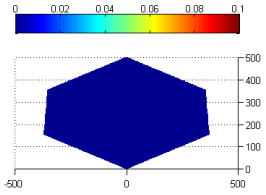
(m) image at 035° rotation



(n) CF precision at 035° rotation

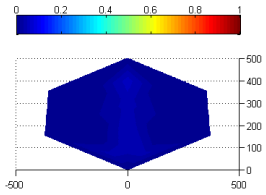


(o) CF recall at 035° rotation

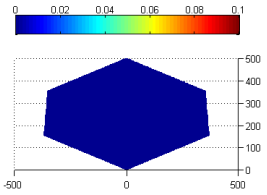


(p) CF efficiency at 035° rotation

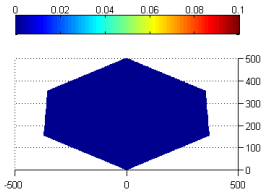
(q) image at 160° rotation



(r) CF precision at 160° rotation

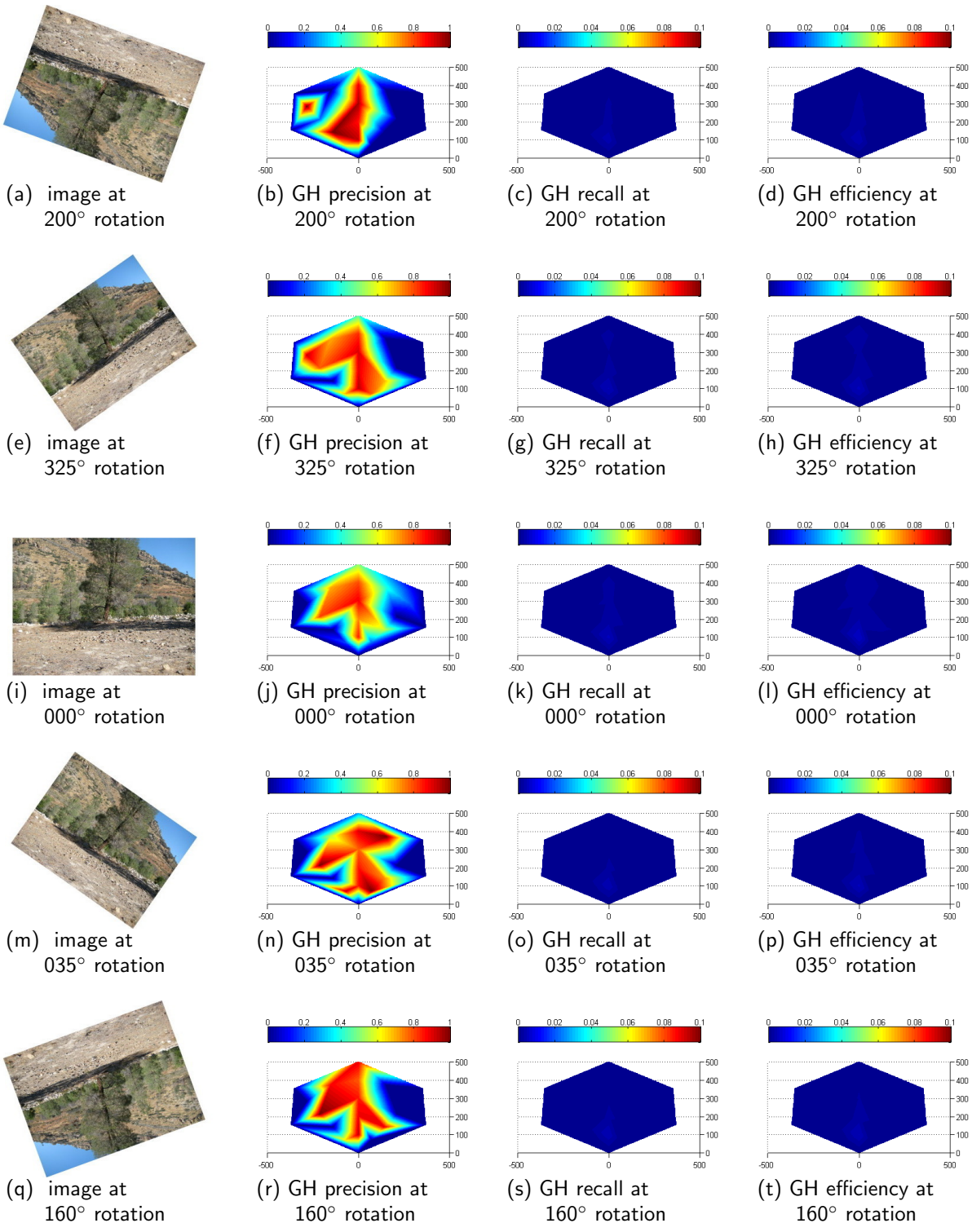


(s) CF recall at 160° rotation



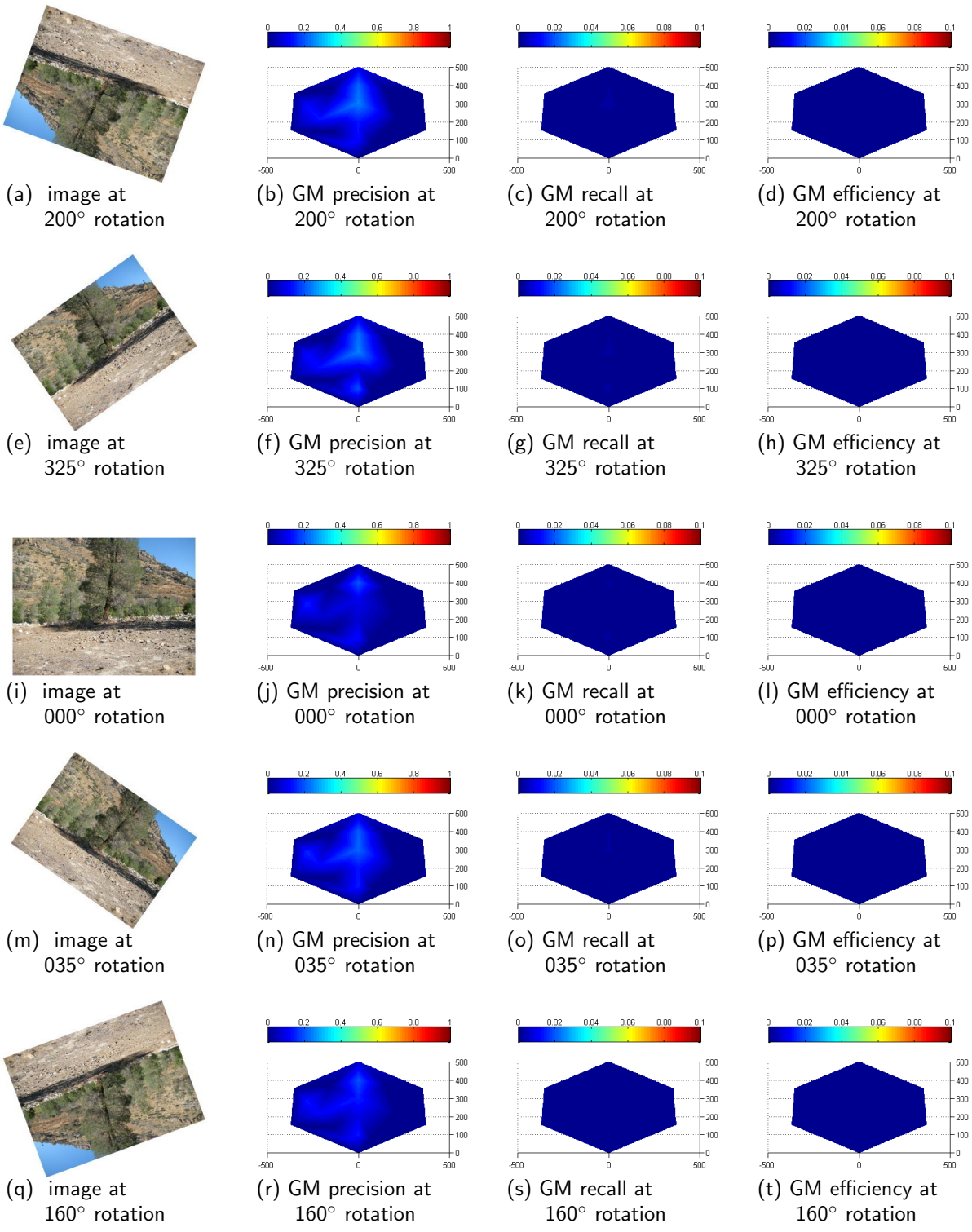
(t) CF efficiency at 160° rotation

Figure 96. Heat maps for descriptor CF in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



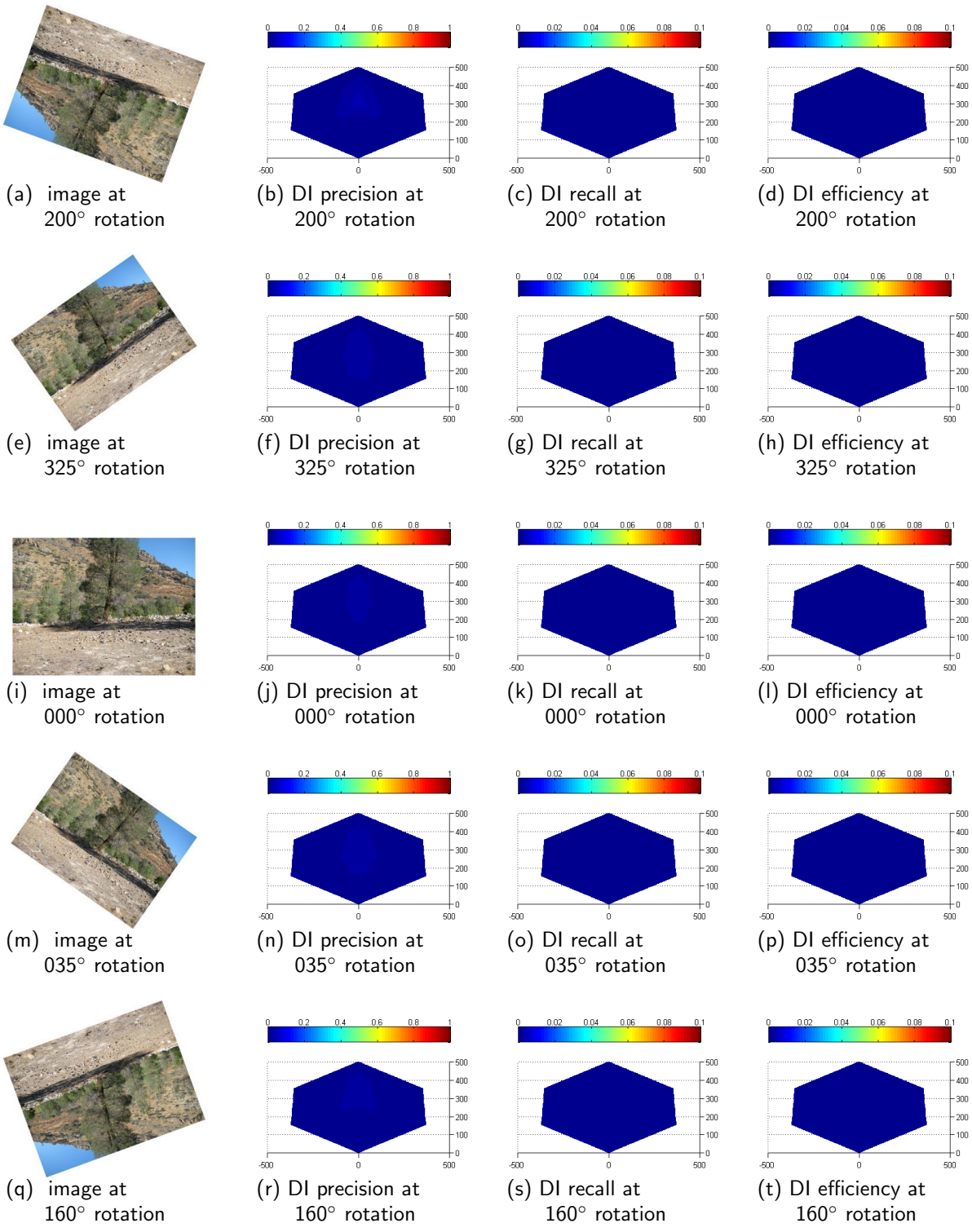
(a) image at 200° rotation
(b) GH precision at 200° rotation
(c) GH recall at 200° rotation
(d) GH efficiency at 200° rotation
(e) image at 325° rotation
(f) GH precision at 325° rotation
(g) GH recall at 325° rotation
(h) GH efficiency at 325° rotation
(i) image at 000° rotation
(j) GH precision at 000° rotation
(k) GH recall at 000° rotation
(l) GH efficiency at 000° rotation
(m) image at 035° rotation
(n) GH precision at 035° rotation
(o) GH recall at 035° rotation
(p) GH efficiency at 035° rotation
(q) image at 160° rotation
(r) GH precision at 160° rotation
(s) GH recall at 160° rotation
(t) GH efficiency at 160° rotation

Figure 97. Heat maps for descriptor GH in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



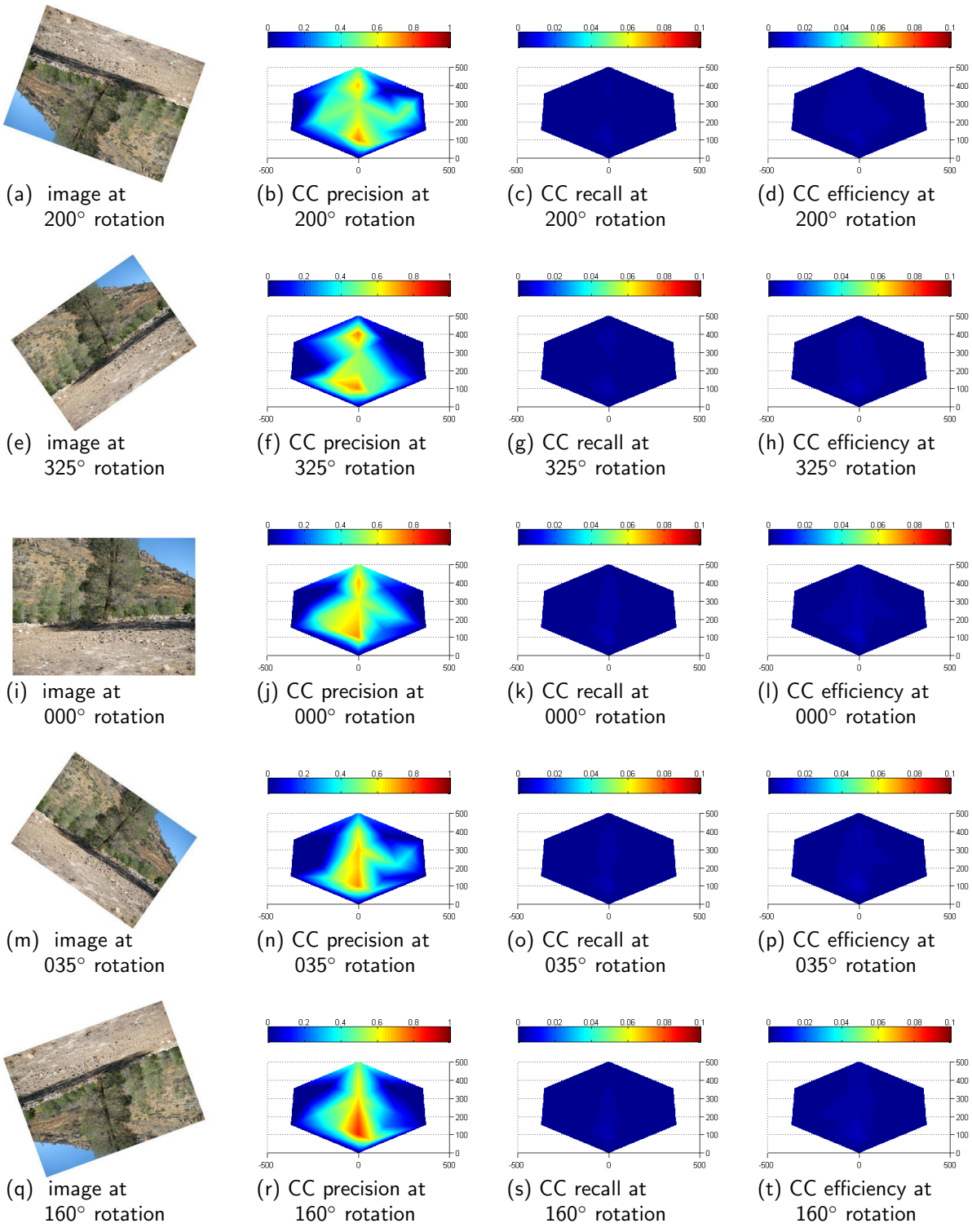
(a) image at 200° rotation
(b) GM precision at 200° rotation
(c) GM recall at 200° rotation
(d) GM efficiency at 200° rotation
(e) image at 325° rotation
(f) GM precision at 325° rotation
(g) GM recall at 325° rotation
(h) GM efficiency at 325° rotation
(i) image at 000° rotation
(j) GM precision at 000° rotation
(k) GM recall at 000° rotation
(l) GM efficiency at 000° rotation
(m) image at 035° rotation
(n) GM precision at 035° rotation
(o) GM recall at 035° rotation
(p) GM efficiency at 035° rotation
(q) image at 160° rotation
(r) GM precision at 160° rotation
(s) GM recall at 160° rotation
(t) GM efficiency at 160° rotation

Figure 98. Heat maps for descriptor GM in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



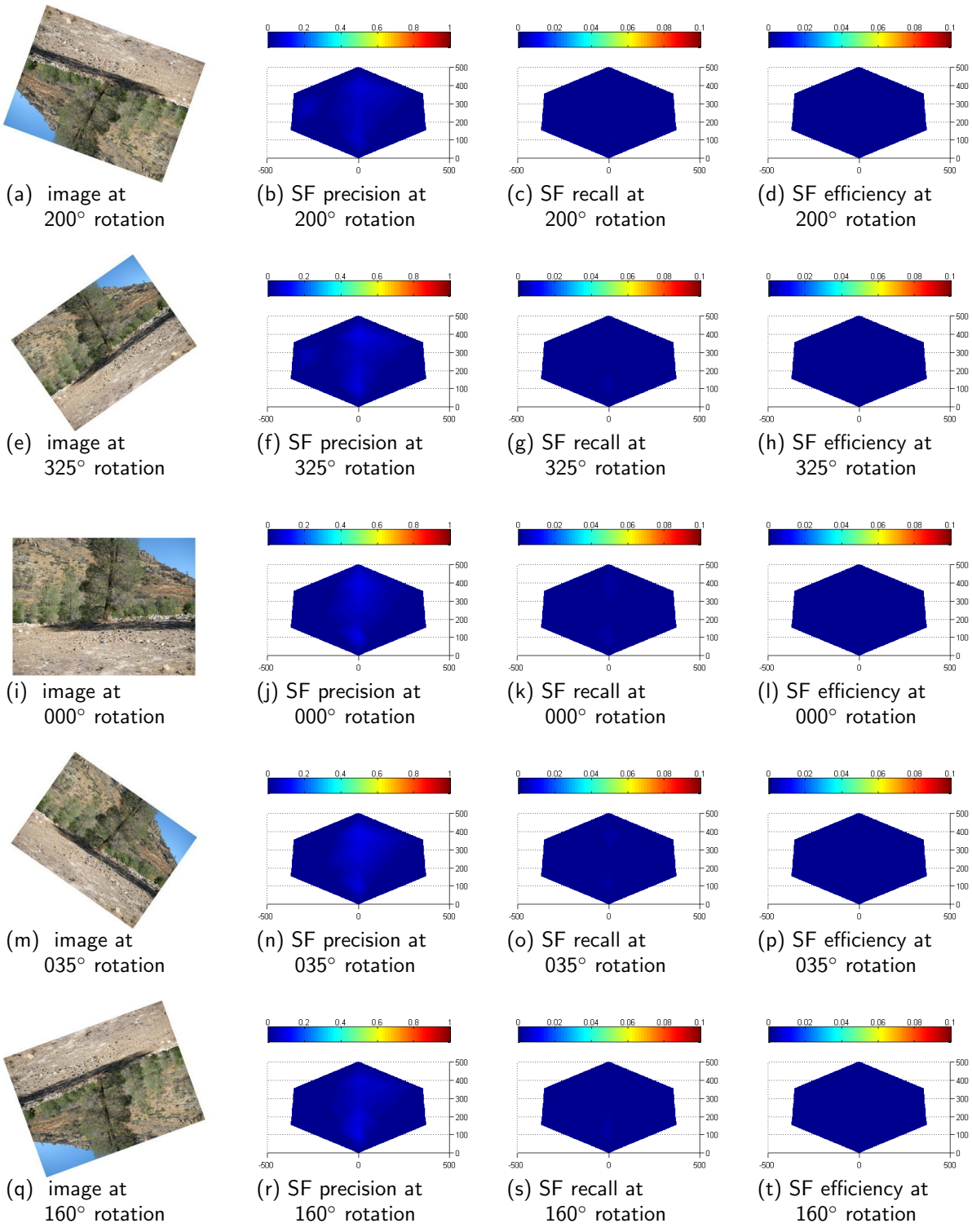
(a) image at 200° rotation
(b) DI precision at 200° rotation
(c) DI recall at 200° rotation
(d) DI efficiency at 200° rotation
(e) image at 325° rotation
(f) DI precision at 325° rotation
(g) DI recall at 325° rotation
(h) DI efficiency at 325° rotation
(i) image at 000° rotation
(j) DI precision at 000° rotation
(k) DI recall at 000° rotation
(l) DI efficiency at 000° rotation
(m) image at 035° rotation
(n) DI precision at 035° rotation
(o) DI recall at 035° rotation
(p) DI efficiency at 035° rotation
(q) image at 160° rotation
(r) DI precision at 160° rotation
(s) DI recall at 160° rotation
(t) DI efficiency at 160° rotation

Figure 99. Heat maps for descriptor DI in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



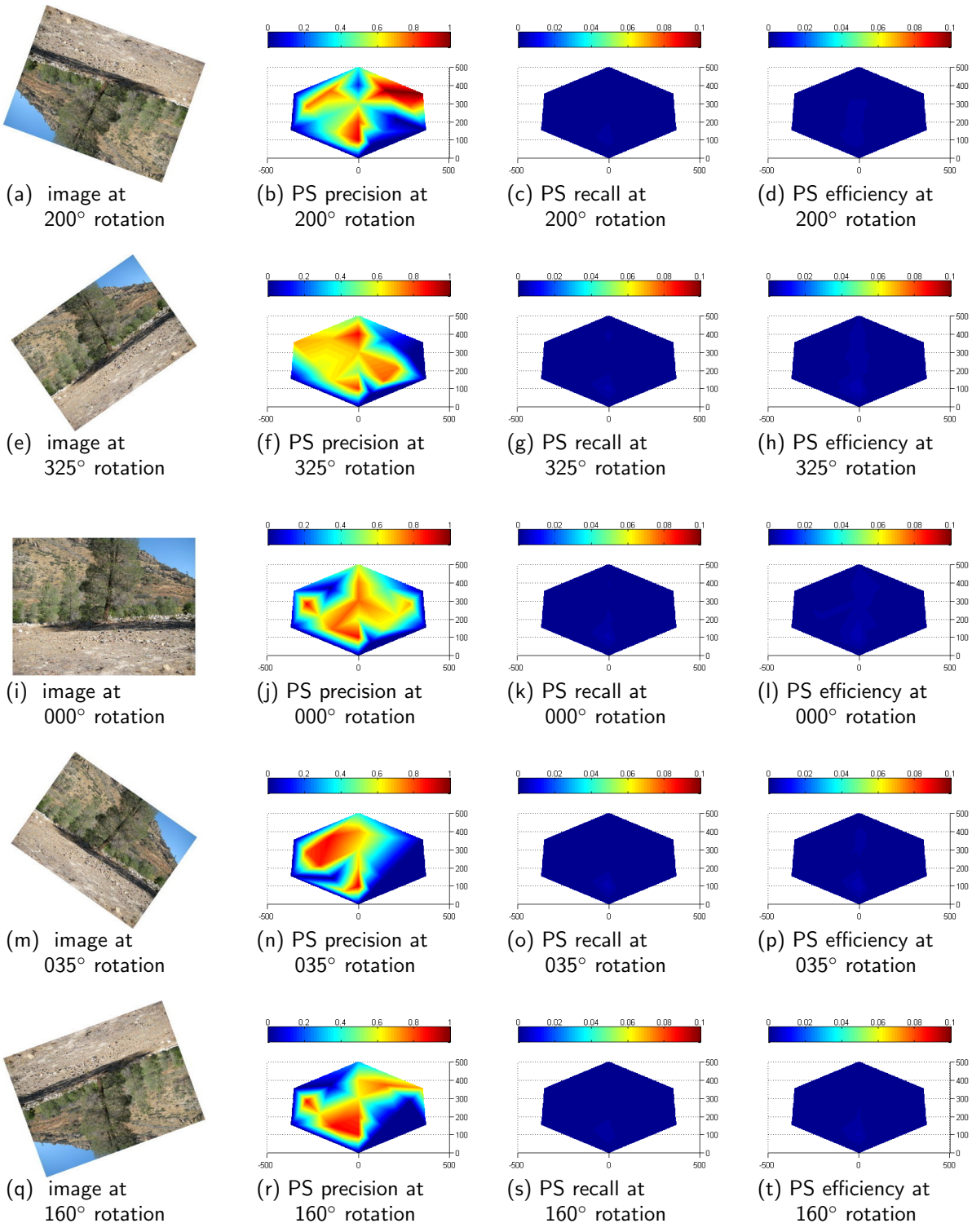
(a) image at 200° rotation
(b) CC precision at 200° rotation
(c) CC recall at 200° rotation
(d) CC efficiency at 200° rotation
(e) image at 325° rotation
(f) CC precision at 325° rotation
(g) CC recall at 325° rotation
(h) CC efficiency at 325° rotation
(i) image at 000° rotation
(j) CC precision at 000° rotation
(k) CC recall at 000° rotation
(l) CC efficiency at 000° rotation
(m) image at 035° rotation
(n) CC precision at 035° rotation
(o) CC recall at 035° rotation
(p) CC efficiency at 035° rotation
(q) image at 160° rotation
(r) CC precision at 160° rotation
(s) CC recall at 160° rotation
(t) CC efficiency at 160° rotation

Figure 100. Heat maps for descriptor CC in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



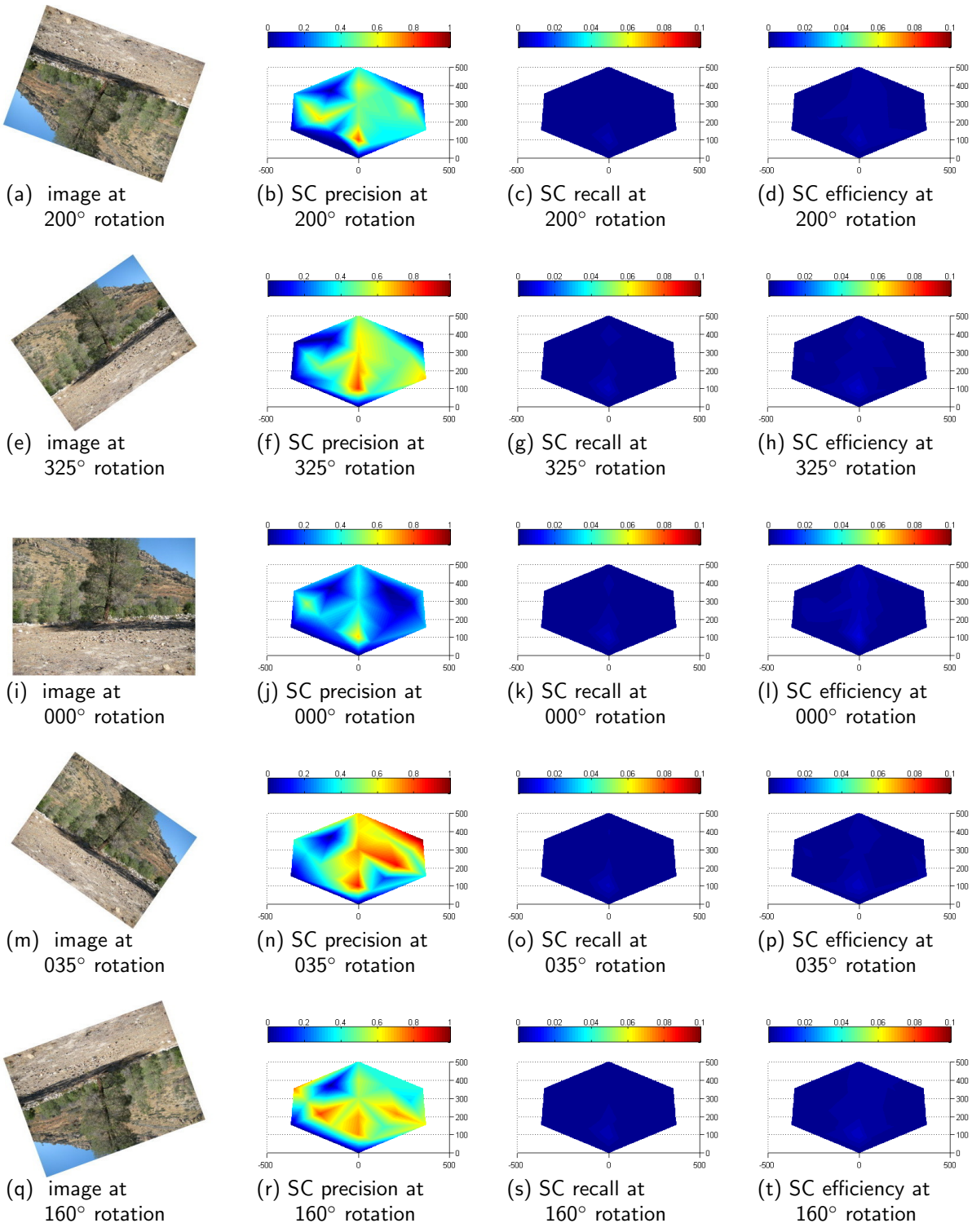
(a) image at 200° rotation
(b) SF precision at 200° rotation
(c) SF recall at 200° rotation
(d) SF efficiency at 200° rotation
(e) image at 325° rotation
(f) SF precision at 325° rotation
(g) SF recall at 325° rotation
(h) SF efficiency at 325° rotation
(i) image at 000° rotation
(j) SF precision at 000° rotation
(k) SF recall at 000° rotation
(l) SF efficiency at 000° rotation
(m) image at 035° rotation
(n) SF precision at 035° rotation
(o) SF recall at 035° rotation
(p) SF efficiency at 035° rotation
(q) image at 160° rotation
(r) SF precision at 160° rotation
(s) SF recall at 160° rotation
(t) SF efficiency at 160° rotation

Figure 101. Heat maps for descriptor SF in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



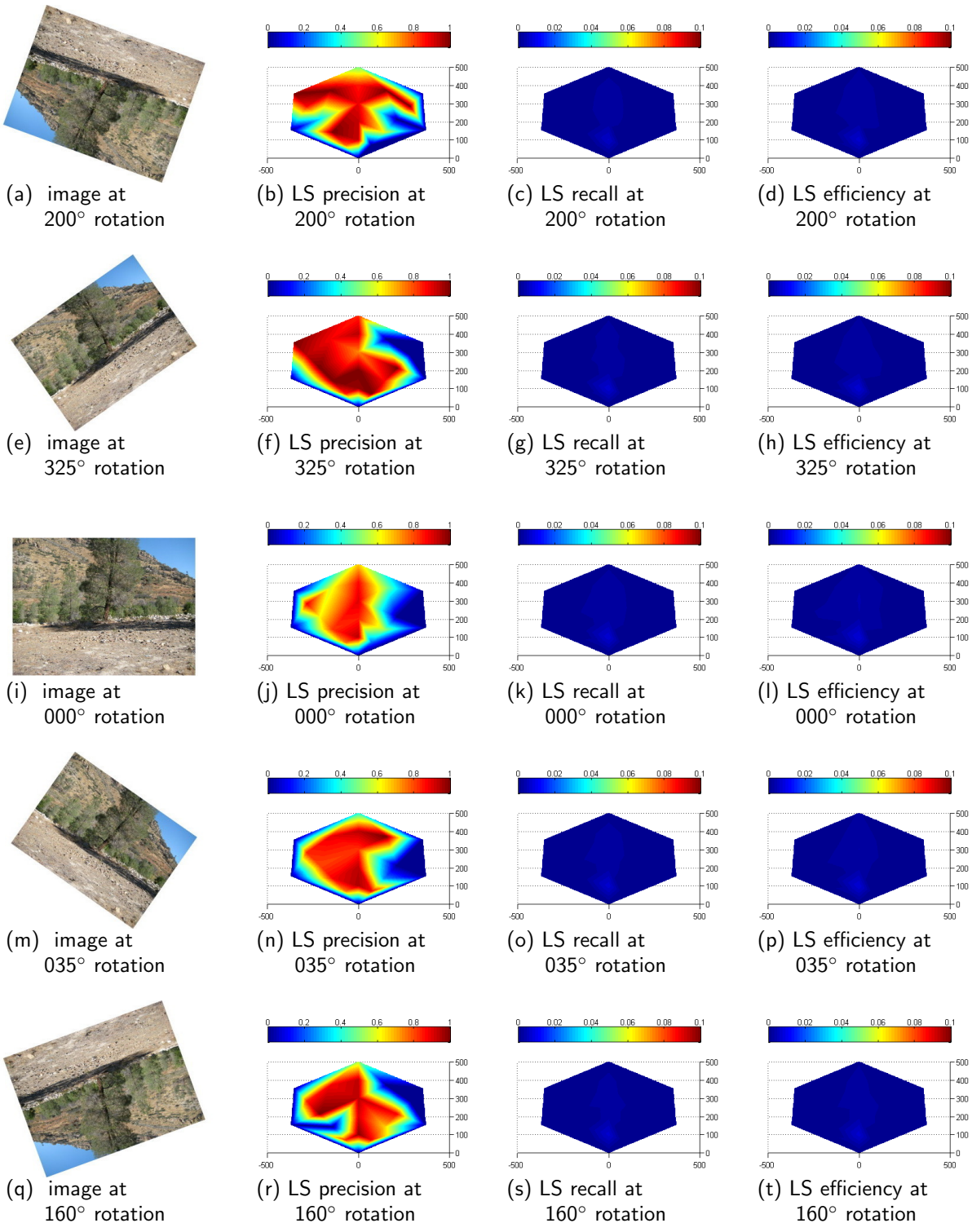
(a) image at 200° rotation
(b) PS precision at 200° rotation
(c) PS recall at 200° rotation
(d) PS efficiency at 200° rotation
(e) image at 325° rotation
(f) PS precision at 325° rotation
(g) PS recall at 325° rotation
(h) PS efficiency at 325° rotation
(i) image at 000° rotation
(j) PS precision at 000° rotation
(k) PS recall at 000° rotation
(l) PS efficiency at 000° rotation
(m) image at 035° rotation
(n) PS precision at 035° rotation
(o) PS recall at 035° rotation
(p) PS efficiency at 035° rotation
(q) image at 160° rotation
(r) PS precision at 160° rotation
(s) PS recall at 160° rotation
(t) PS efficiency at 160° rotation

Figure 102. Heat maps for descriptor PS in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



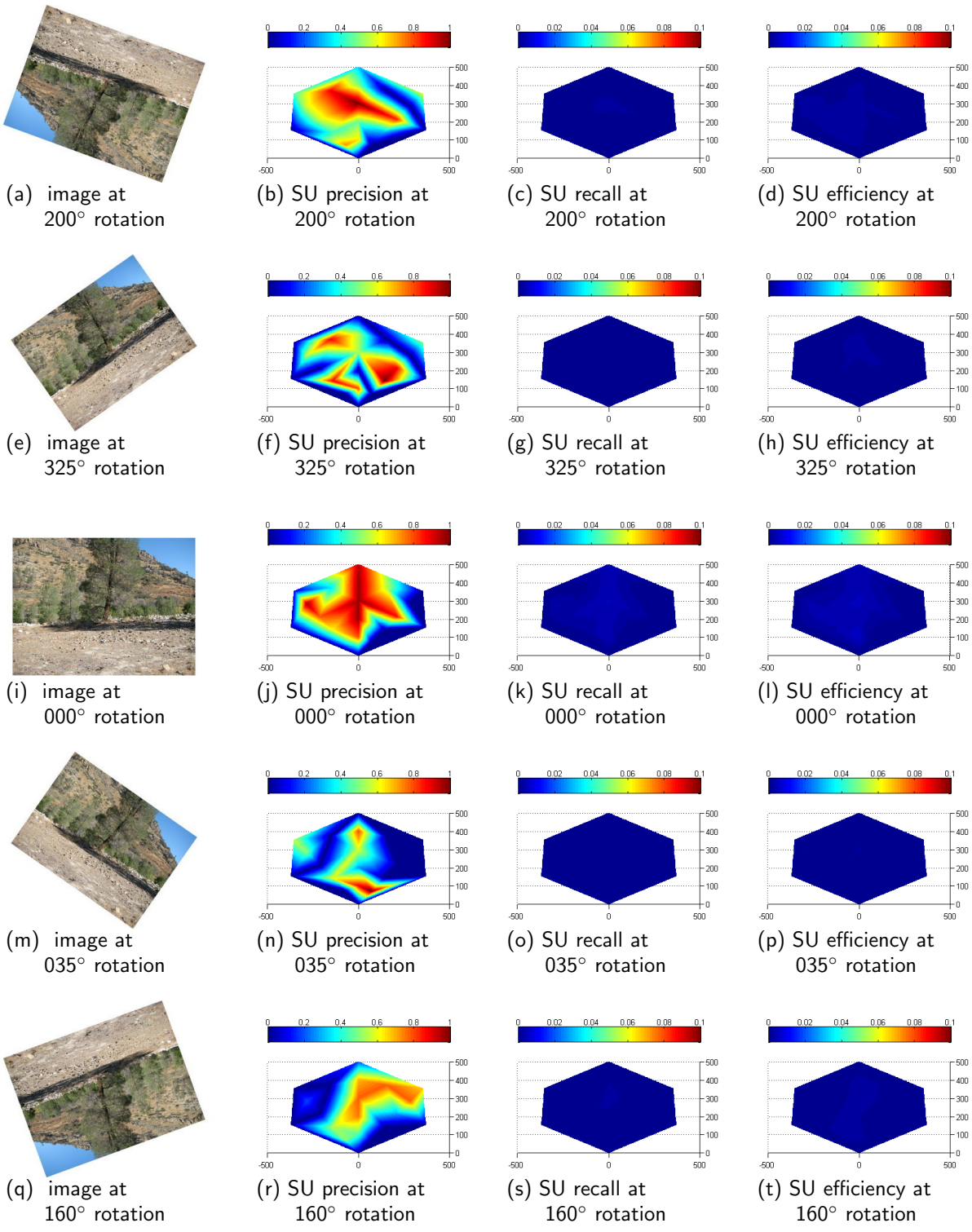
(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

Figure 103. Heat maps for descriptor SC in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



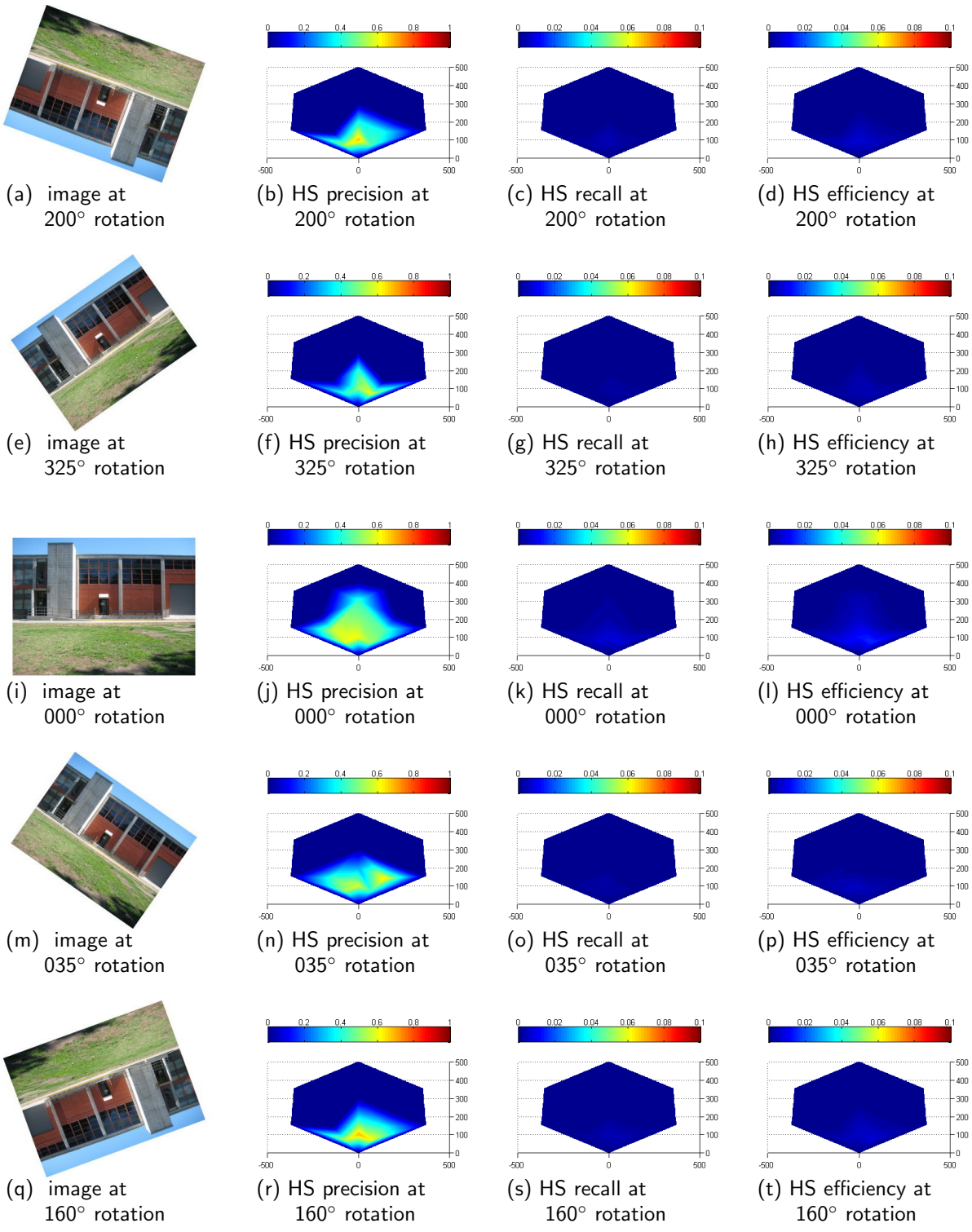
(a) image at 200° rotation
(b) LS precision at 200° rotation
(c) LS recall at 200° rotation
(d) LS efficiency at 200° rotation
(e) image at 325° rotation
(f) LS precision at 325° rotation
(g) LS recall at 325° rotation
(h) LS efficiency at 325° rotation
(i) image at 000° rotation
(j) LS precision at 000° rotation
(k) LS recall at 000° rotation
(l) LS efficiency at 000° rotation
(m) image at 035° rotation
(n) LS precision at 035° rotation
(o) LS recall at 035° rotation
(p) LS efficiency at 035° rotation
(q) image at 160° rotation
(r) LS precision at 160° rotation
(s) LS recall at 160° rotation
(t) LS efficiency at 160° rotation

Figure 104. Heat maps for descriptor LS in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) SU precision at 200° rotation
(c) SU recall at 200° rotation
(d) SU efficiency at 200° rotation
(e) image at 325° rotation
(f) SU precision at 325° rotation
(g) SU recall at 325° rotation
(h) SU efficiency at 325° rotation
(i) image at 000° rotation
(j) SU precision at 000° rotation
(k) SU recall at 000° rotation
(l) SU efficiency at 000° rotation
(m) image at 035° rotation
(n) SU precision at 035° rotation
(o) SU recall at 035° rotation
(p) SU efficiency at 035° rotation
(q) image at 160° rotation
(r) SU precision at 160° rotation
(s) SU recall at 160° rotation
(t) SU efficiency at 160° rotation

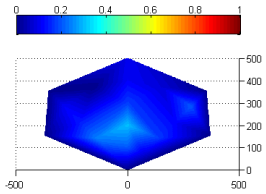
Figure 105. Heat maps for descriptor SU in the OutTree scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



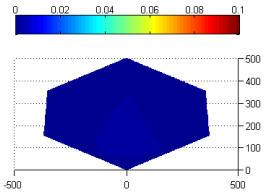
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 106. Heat maps for descriptor HS in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

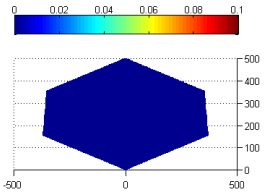
(a) image at 200° rotation



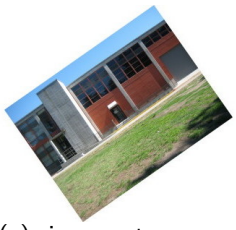
(b) CF precision at 200° rotation



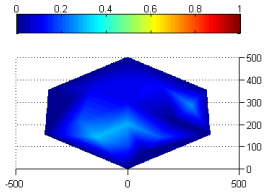
(c) CF recall at 200° rotation



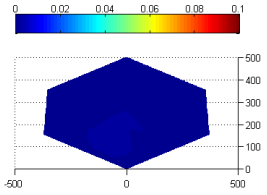
(d) CF efficiency at 200° rotation



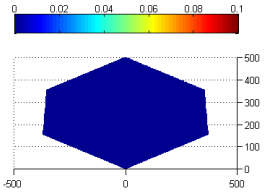
(e) image at 325° rotation



(f) CF precision at 325° rotation

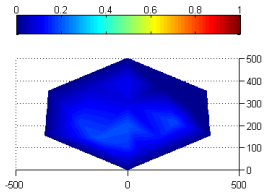


(g) CF recall at 325° rotation

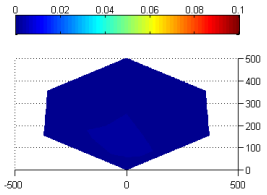


(h) CF efficiency at 325° rotation

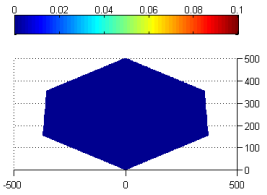
(i) image at 000° rotation



(j) CF precision at 000° rotation

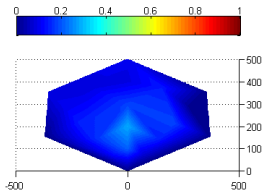


(k) CF recall at 000° rotation

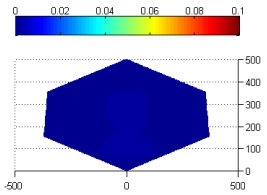


(l) CF efficiency at 000° rotation

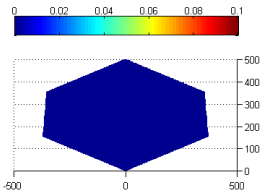
(m) image at 035° rotation



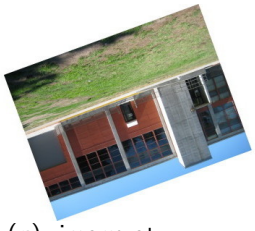
(n) CF precision at 035° rotation



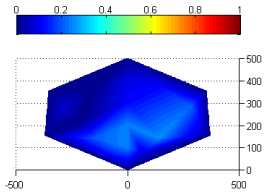
(o) CF recall at 035° rotation



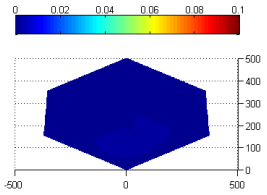
(p) CF efficiency at 035° rotation



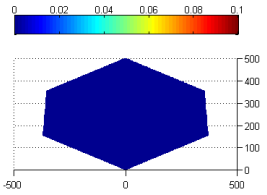
(q) image at 160° rotation



(r) CF precision at 160° rotation

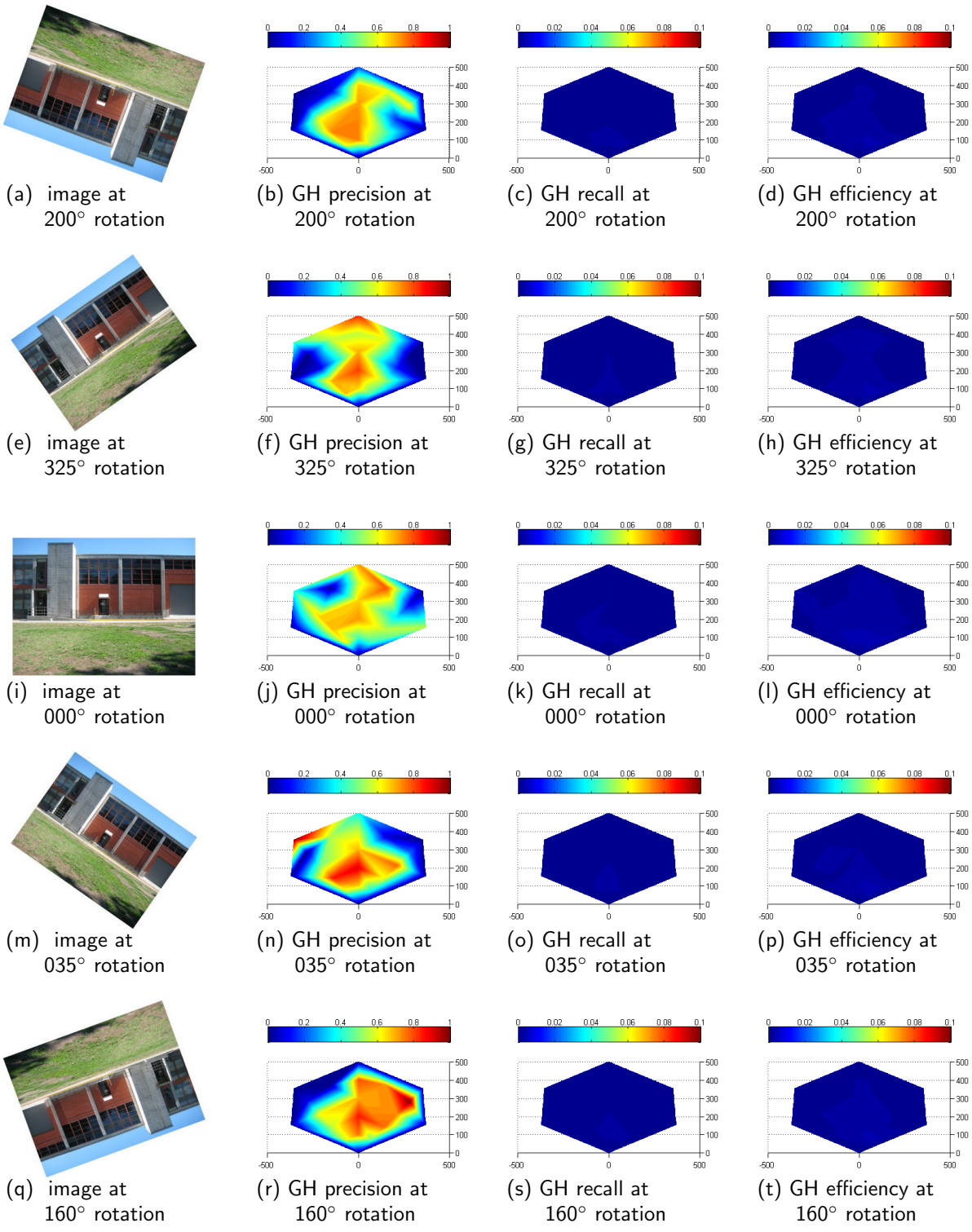


(s) CF recall at 160° rotation



(t) CF efficiency at 160° rotation

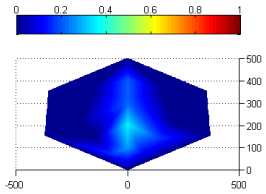
Figure 107. Heat maps for descriptor CF in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



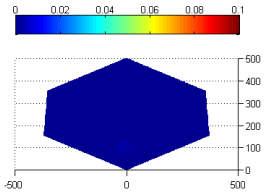
(a) image at 200° rotation
(b) GH precision at 200° rotation
(c) GH recall at 200° rotation
(d) GH efficiency at 200° rotation
(e) image at 325° rotation
(f) GH precision at 325° rotation
(g) GH recall at 325° rotation
(h) GH efficiency at 325° rotation
(i) image at 000° rotation
(j) GH precision at 000° rotation
(k) GH recall at 000° rotation
(l) GH efficiency at 000° rotation
(m) image at 035° rotation
(n) GH precision at 035° rotation
(o) GH recall at 035° rotation
(p) GH efficiency at 035° rotation
(q) image at 160° rotation
(r) GH precision at 160° rotation
(s) GH recall at 160° rotation
(t) GH efficiency at 160° rotation

Figure 108. Heat maps for descriptor GH in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

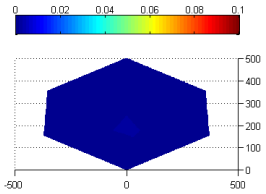
(a) image at 200° rotation



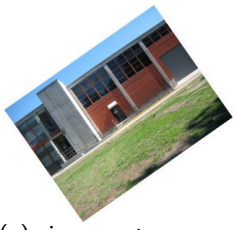
(b) GM precision at 200° rotation



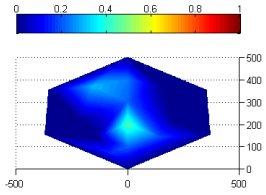
(c) GM recall at 200° rotation



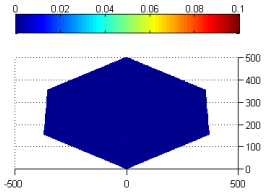
(d) GM efficiency at 200° rotation



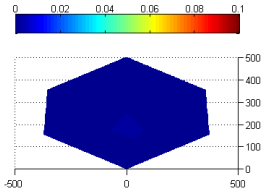
(e) image at 325° rotation



(f) GM precision at 325° rotation

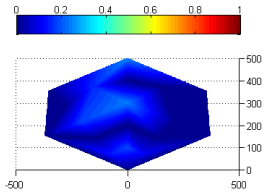


(g) GM recall at 325° rotation

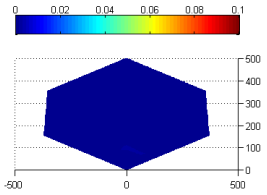


(h) GM efficiency at 325° rotation

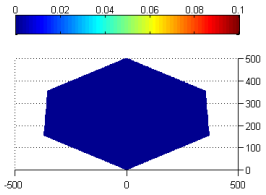
(i) image at 000° rotation



(j) GM precision at 000° rotation

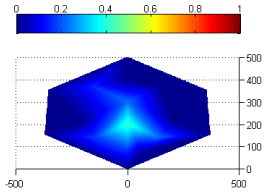


(k) GM recall at 000° rotation

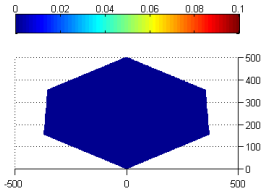


(l) GM efficiency at 000° rotation

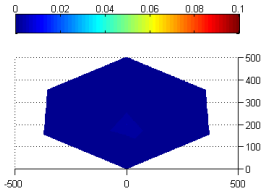
(m) image at 035° rotation



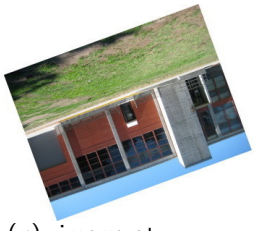
(n) GM precision at 035° rotation



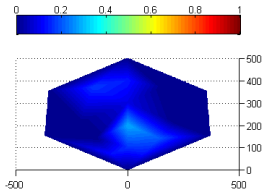
(o) GM recall at 035° rotation



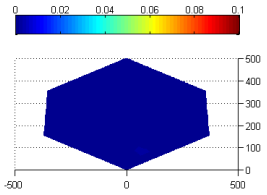
(p) GM efficiency at 035° rotation



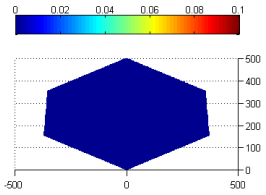
(q) image at 160° rotation



(r) GM precision at 160° rotation



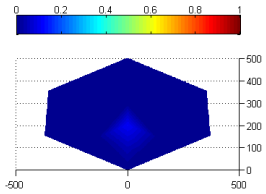
(s) GM recall at 160° rotation



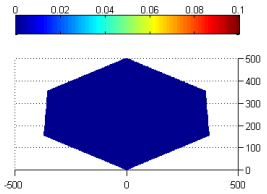
(t) GM efficiency at 160° rotation

Figure 109. Heat maps for descriptor GM in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

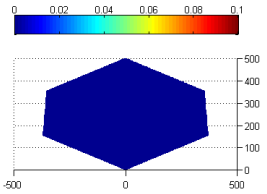
(a) image at 200° rotation



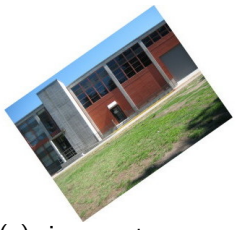
(b) DI precision at 200° rotation



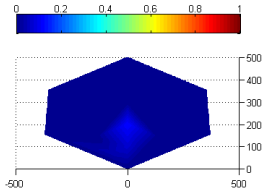
(c) DI recall at 200° rotation



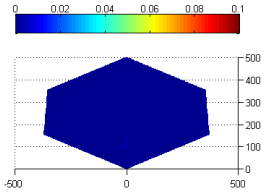
(d) DI efficiency at 200° rotation



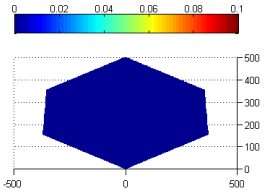
(e) image at 325° rotation



(f) DI precision at 325° rotation

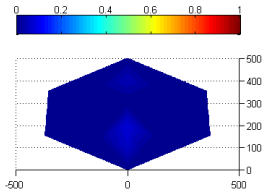


(g) DI recall at 325° rotation

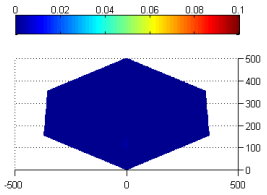


(h) DI efficiency at 325° rotation

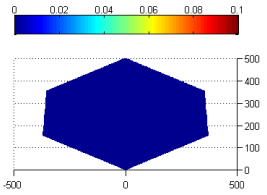
(i) image at 000° rotation



(j) DI precision at 000° rotation

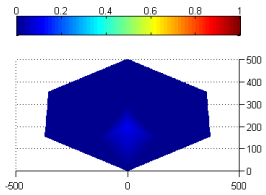


(k) DI recall at 000° rotation

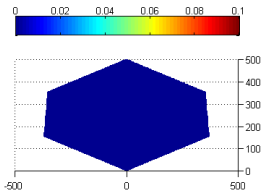


(l) DI efficiency at 000° rotation

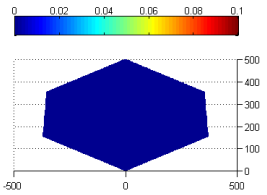
(m) image at 035° rotation



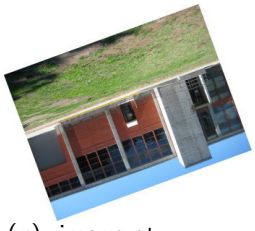
(n) DI precision at 035° rotation



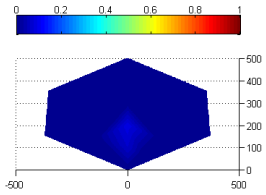
(o) DI recall at 035° rotation



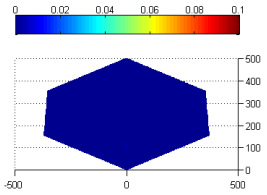
(p) DI efficiency at 035° rotation



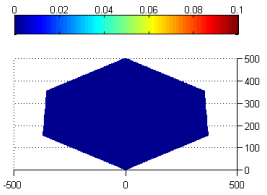
(q) image at 160° rotation



(r) DI precision at 160° rotation

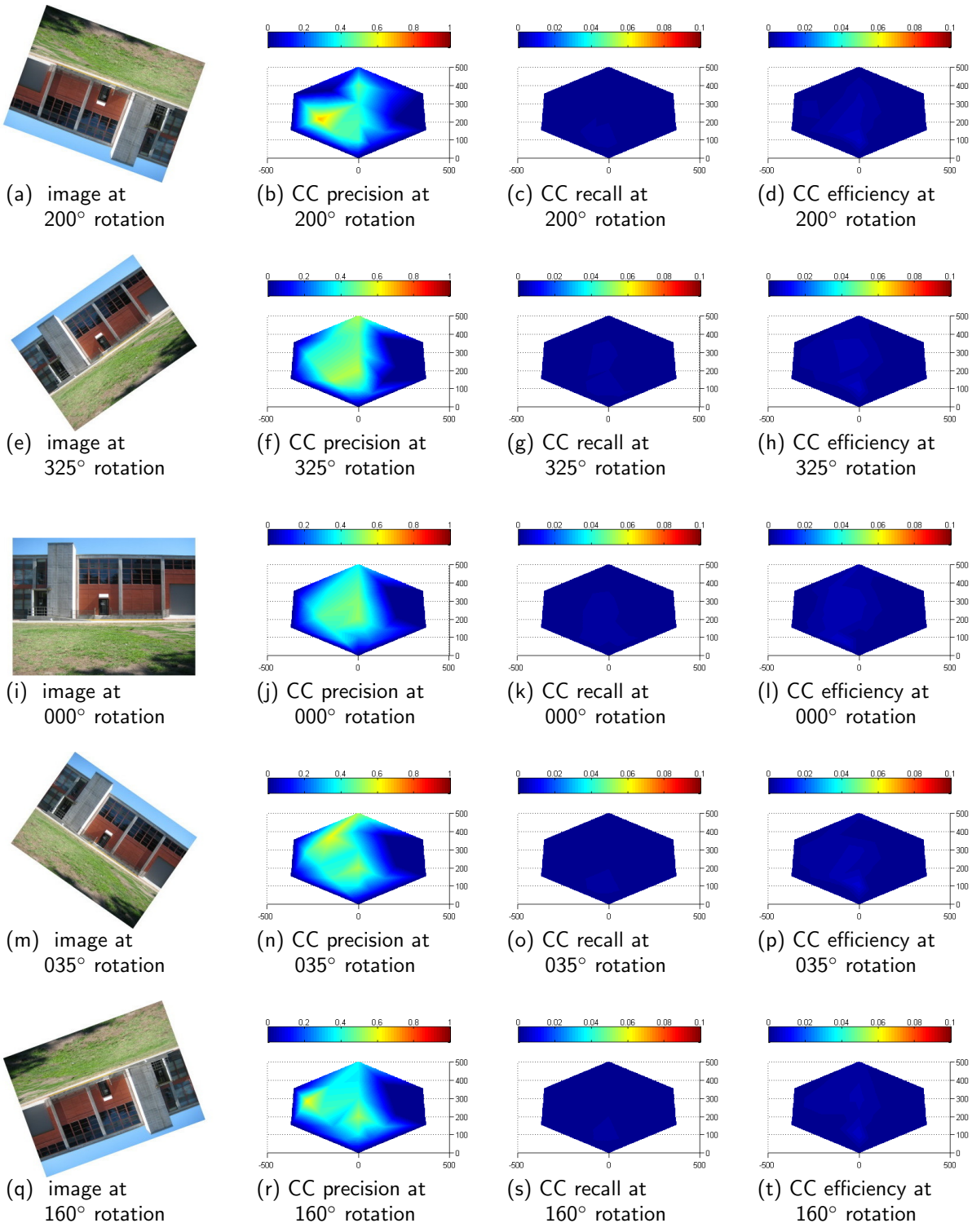


(s) DI recall at 160° rotation



(t) DI efficiency at 160° rotation

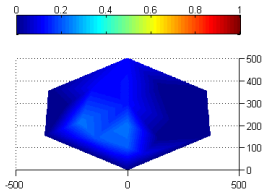
Figure 110. Heat maps for descriptor DI in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



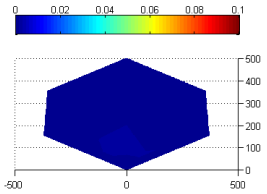
(a) image at 200° rotation
(b) CC precision at 200° rotation
(c) CC recall at 200° rotation
(d) CC efficiency at 200° rotation
(e) image at 325° rotation
(f) CC precision at 325° rotation
(g) CC recall at 325° rotation
(h) CC efficiency at 325° rotation
(i) image at 000° rotation
(j) CC precision at 000° rotation
(k) CC recall at 000° rotation
(l) CC efficiency at 000° rotation
(m) image at 035° rotation
(n) CC precision at 035° rotation
(o) CC recall at 035° rotation
(p) CC efficiency at 035° rotation
(q) image at 160° rotation
(r) CC precision at 160° rotation
(s) CC recall at 160° rotation
(t) CC efficiency at 160° rotation

Figure 111. Heat maps for descriptor CC in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

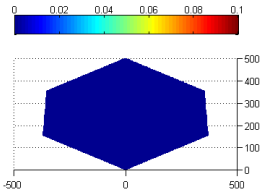
(a) image at 200° rotation



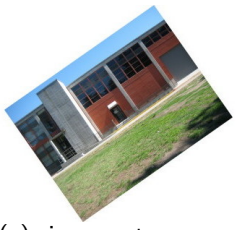
(b) SF precision at 200° rotation



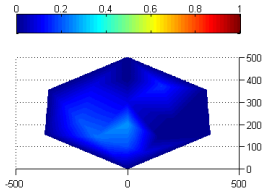
(c) SF recall at 200° rotation



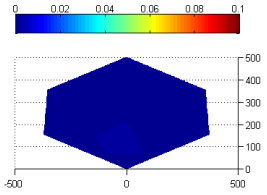
(d) SF efficiency at 200° rotation



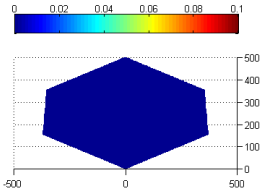
(e) image at 325° rotation



(f) SF precision at 325° rotation

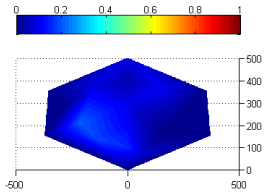


(g) SF recall at 325° rotation

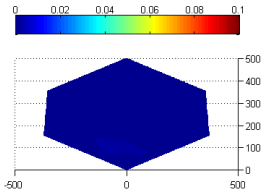


(h) SF efficiency at 325° rotation

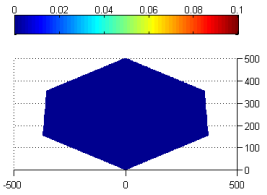
(i) image at 000° rotation



(j) SF precision at 000° rotation

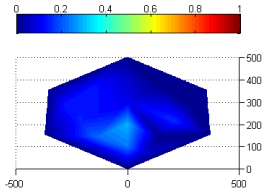


(k) SF recall at 000° rotation

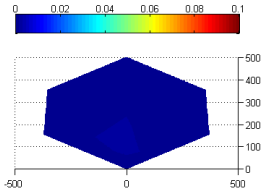


(l) SF efficiency at 000° rotation

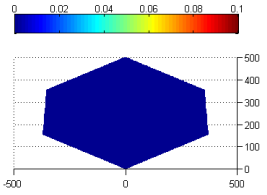
(m) image at 035° rotation



(n) SF precision at 035° rotation

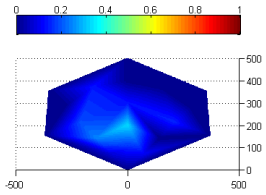


(o) SF recall at 035° rotation

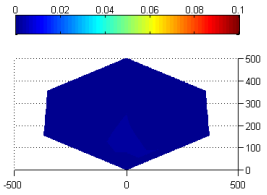


(p) SF efficiency at 035° rotation

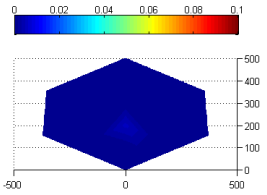
(q) image at 160° rotation



(r) SF precision at 160° rotation

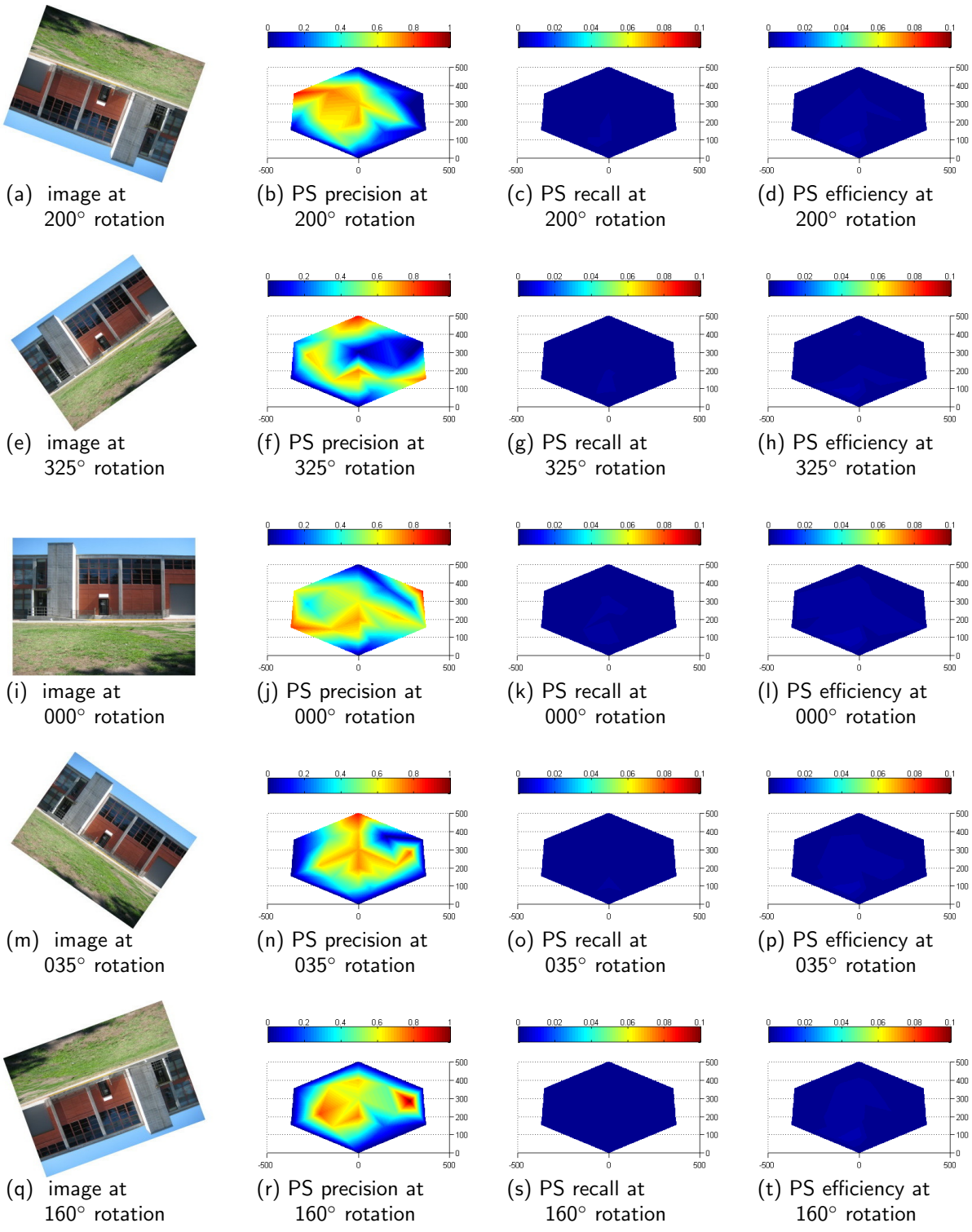


(s) SF recall at 160° rotation



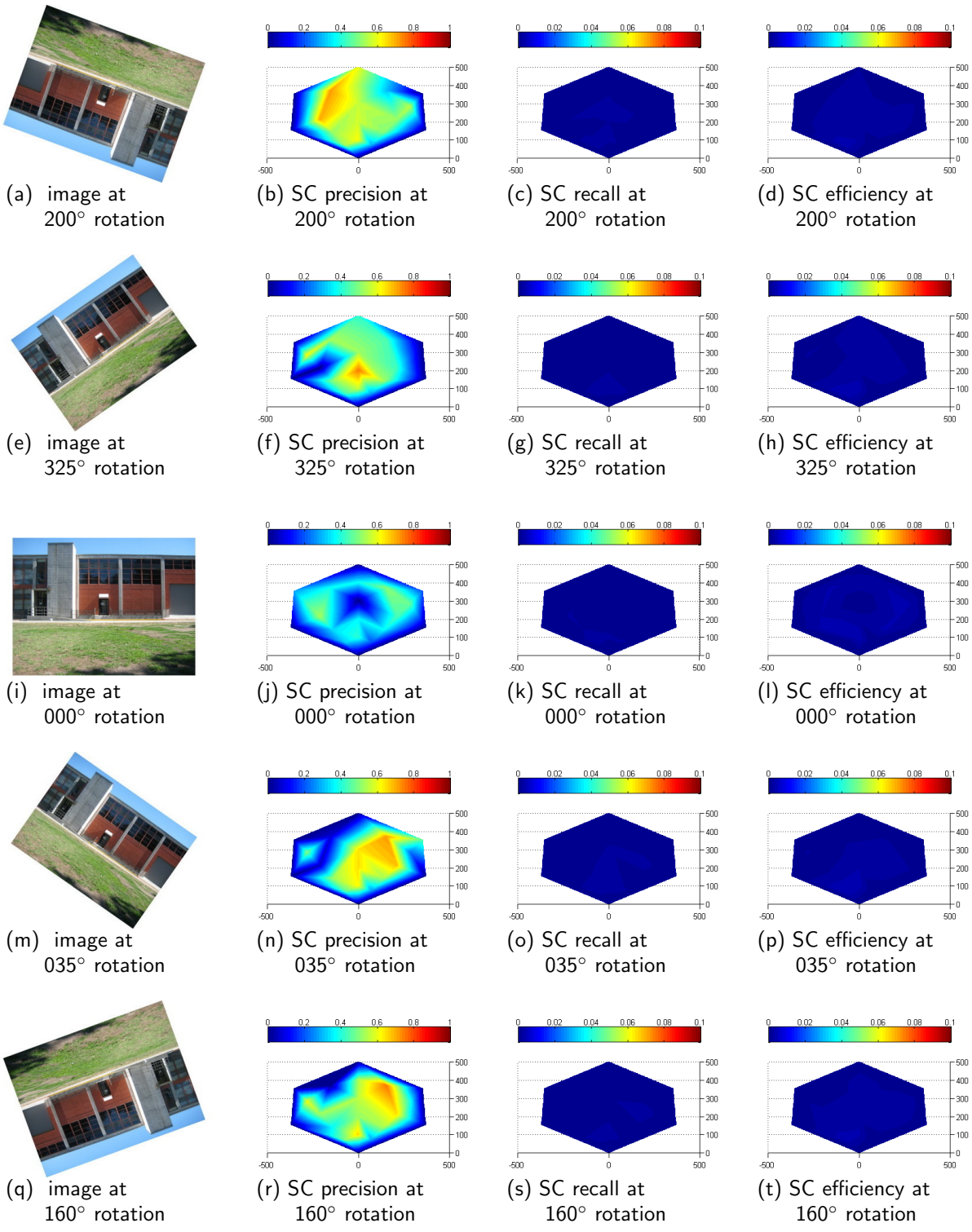
(t) SF efficiency at 160° rotation

Figure 112. Heat maps for descriptor SF in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) PS precision at 200° rotation
(c) PS recall at 200° rotation
(d) PS efficiency at 200° rotation
(e) image at 325° rotation
(f) PS precision at 325° rotation
(g) PS recall at 325° rotation
(h) PS efficiency at 325° rotation
(i) image at 000° rotation
(j) PS precision at 000° rotation
(k) PS recall at 000° rotation
(l) PS efficiency at 000° rotation
(m) image at 035° rotation
(n) PS precision at 035° rotation
(o) PS recall at 035° rotation
(p) PS efficiency at 035° rotation
(q) image at 160° rotation
(r) PS precision at 160° rotation
(s) PS recall at 160° rotation
(t) PS efficiency at 160° rotation

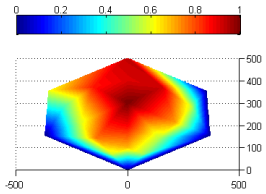
Figure 113. Heat maps for descriptor PS in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



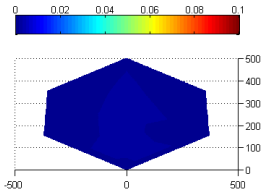
(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

Figure 114. Heat maps for descriptor SC in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.

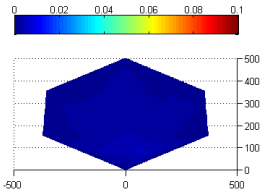
(a) image at 200° rotation



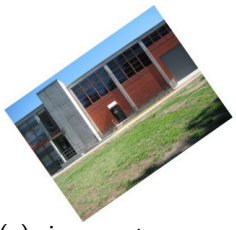
(b) LS precision at 200° rotation



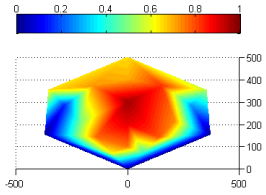
(c) LS recall at 200° rotation



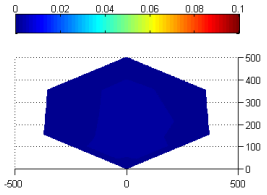
(d) LS efficiency at 200° rotation



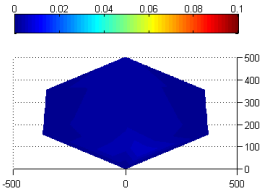
(e) image at 325° rotation



(f) LS precision at 325° rotation

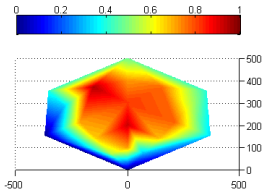


(g) LS recall at 325° rotation

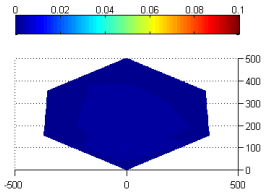


(h) LS efficiency at 325° rotation

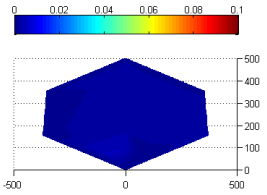
(i) image at 000° rotation



(j) LS precision at 000° rotation

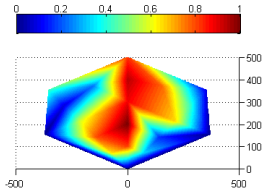


(k) LS recall at 000° rotation

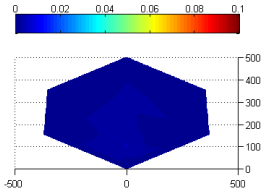


(l) LS efficiency at 000° rotation

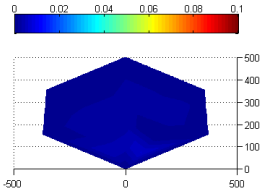
(m) image at 035° rotation



(n) LS precision at 035° rotation

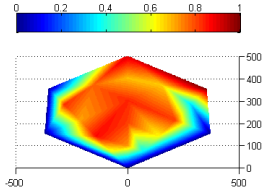


(o) LS recall at 035° rotation

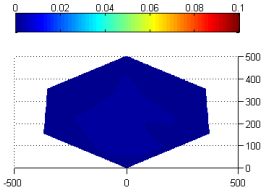


(p) LS efficiency at 035° rotation

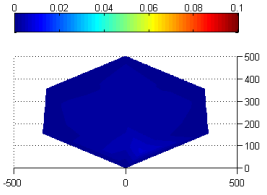
(q) image at 160° rotation



(r) LS precision at 160° rotation

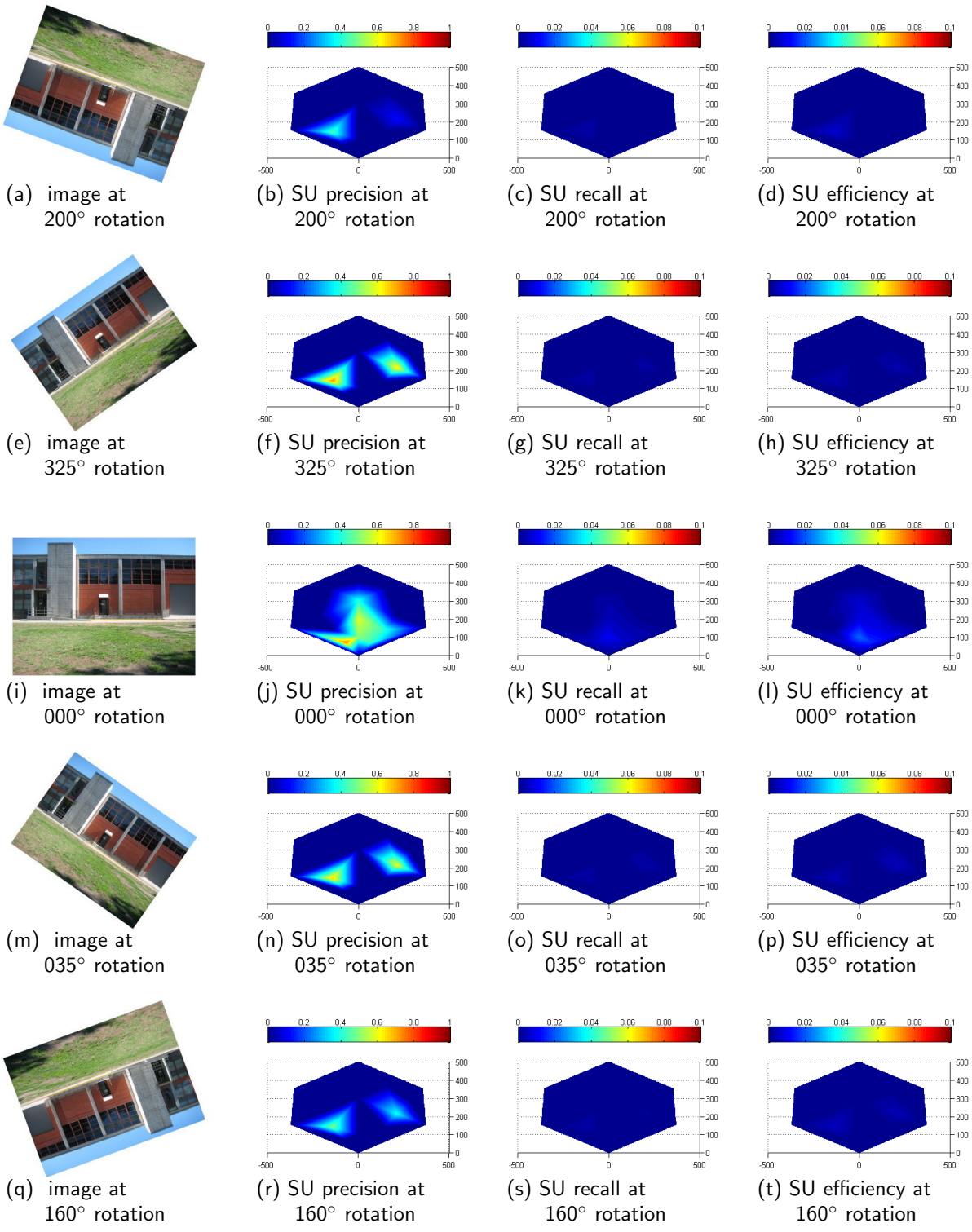


(s) LS recall at 160° rotation



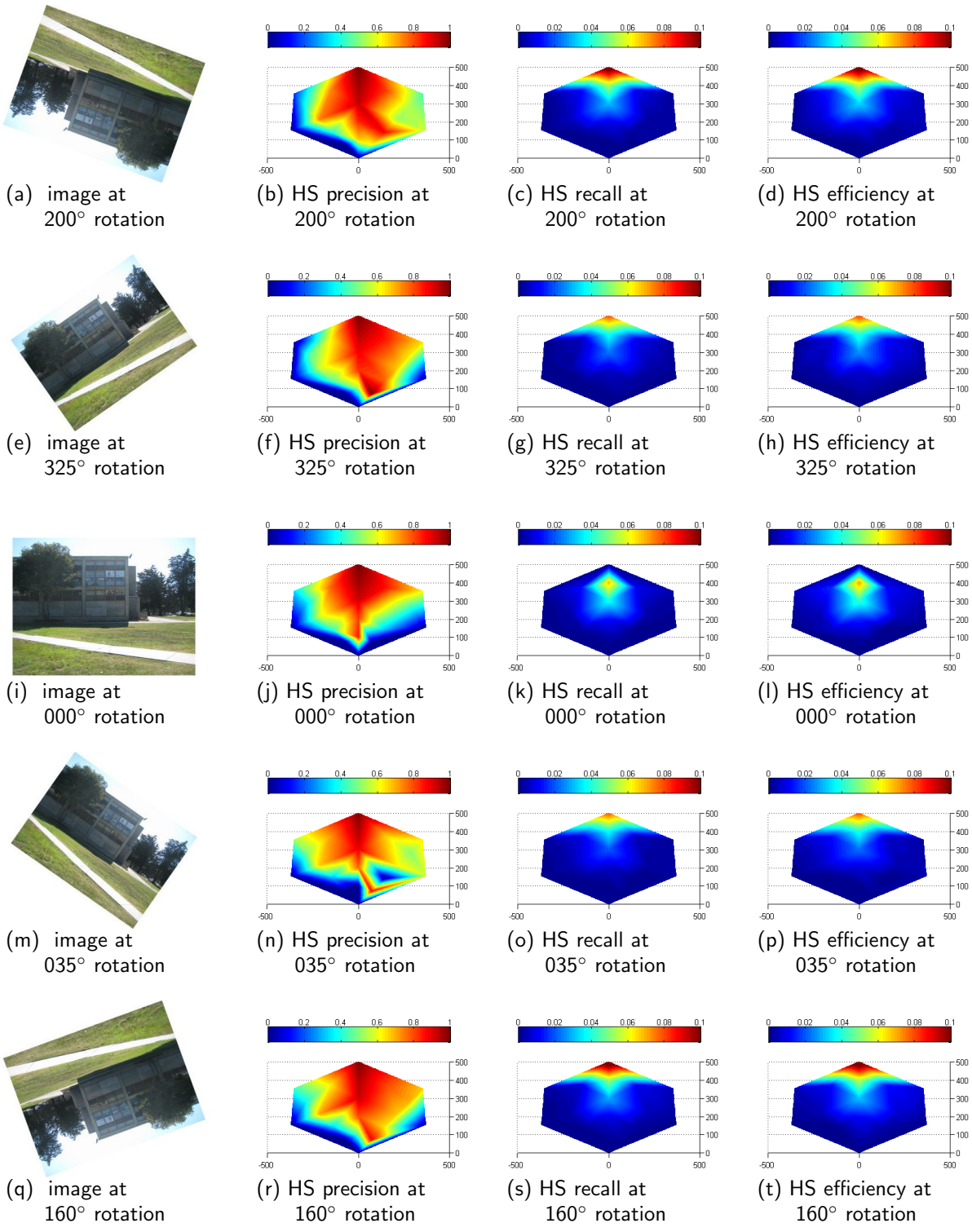
(t) LS efficiency at 160° rotation

Figure 115. Heat maps for descriptor LS in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



(a) image at 200° rotation
(b) SU precision at 200° rotation
(c) SU recall at 200° rotation
(d) SU efficiency at 200° rotation
(e) image at 325° rotation
(f) SU precision at 325° rotation
(g) SU recall at 325° rotation
(h) SU efficiency at 325° rotation
(i) image at 000° rotation
(j) SU precision at 000° rotation
(k) SU recall at 000° rotation
(l) SU efficiency at 000° rotation
(m) image at 035° rotation
(n) SU precision at 035° rotation
(o) SU recall at 035° rotation
(p) SU efficiency at 035° rotation
(q) image at 160° rotation
(r) SU precision at 160° rotation
(s) SU recall at 160° rotation
(t) SU efficiency at 160° rotation

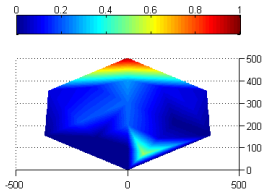
Figure 116. Heat maps for descriptor SU in the OutUSL scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,100). The axis scale is 100=4m.



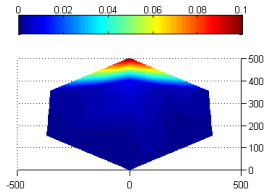
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 117. Heat maps for descriptor HS in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

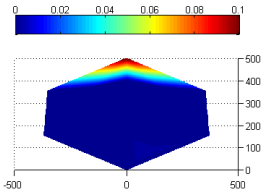
(a) image at 200° rotation



(b) CF precision at 200° rotation

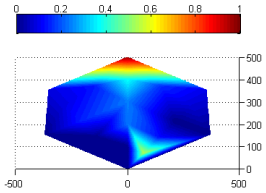


(c) CF recall at 200° rotation

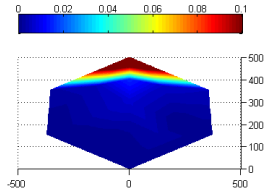


(d) CF efficiency at 200° rotation

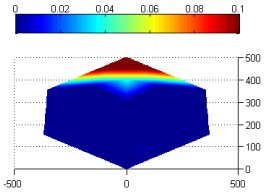
(e) image at 325° rotation



(f) CF precision at 325° rotation

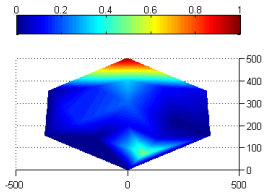


(g) CF recall at 325° rotation

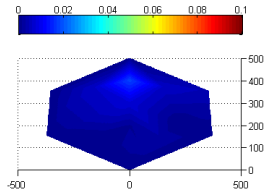


(h) CF efficiency at 325° rotation

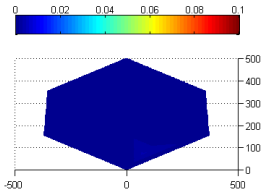
(i) image at 000° rotation



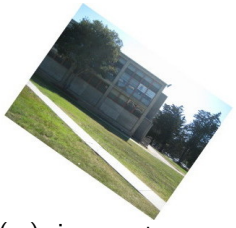
(j) CF precision at 000° rotation



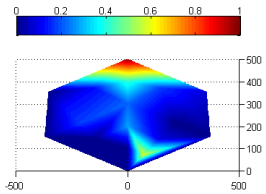
(k) CF recall at 000° rotation



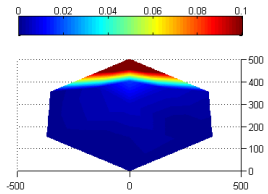
(l) CF efficiency at 000° rotation



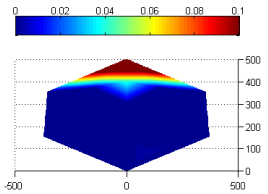
(m) image at 035° rotation



(n) CF precision at 035° rotation

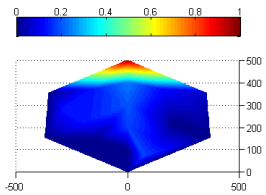


(o) CF recall at 035° rotation

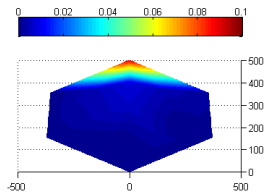


(p) CF efficiency at 035° rotation

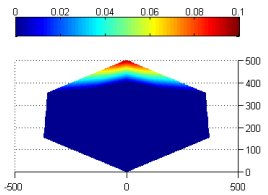
(q) image at 160° rotation



(r) CF precision at 160° rotation

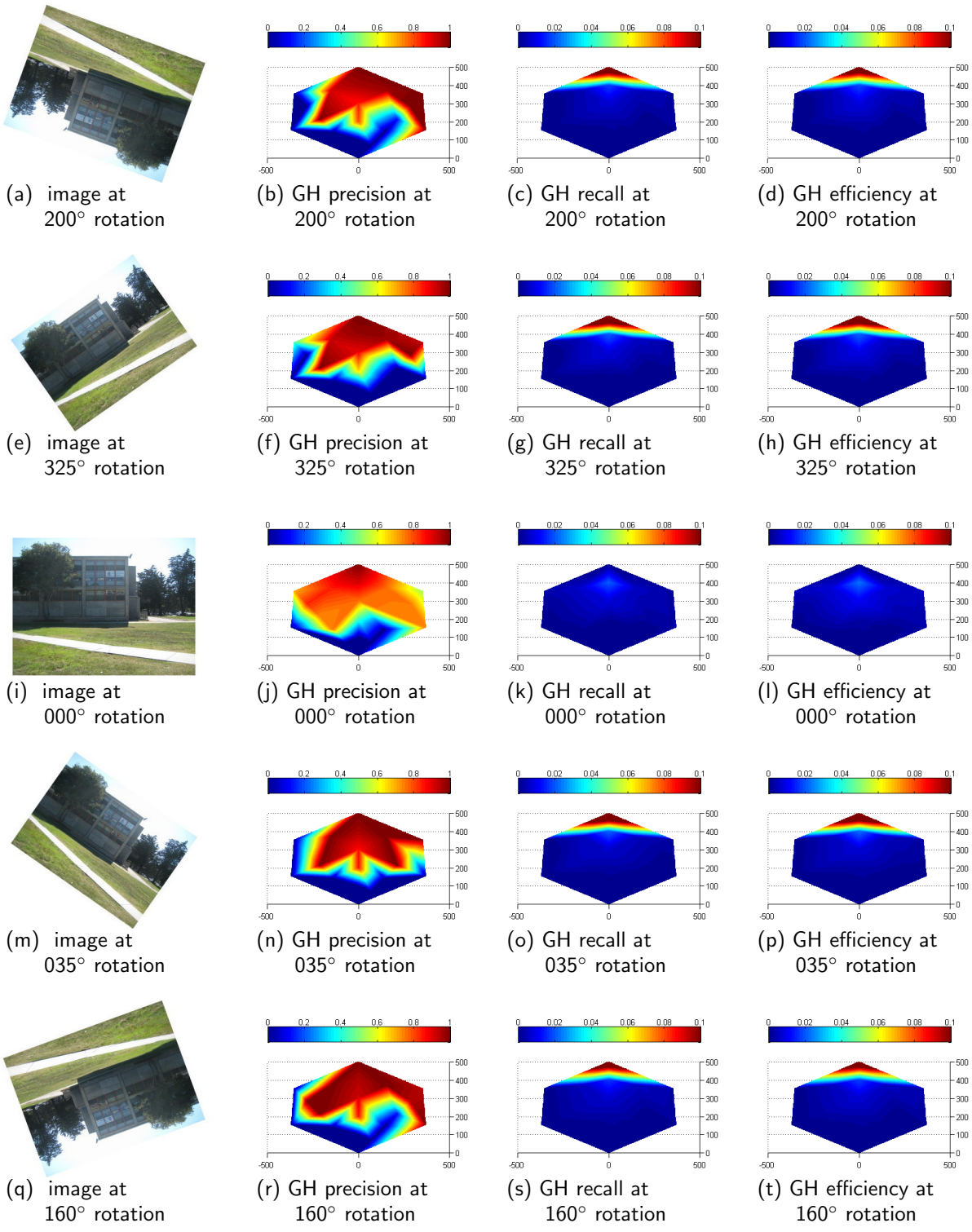


(s) CF recall at 160° rotation



(t) CF efficiency at 160° rotation

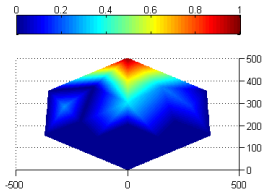
Figure 118. Heat maps for descriptor CF in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



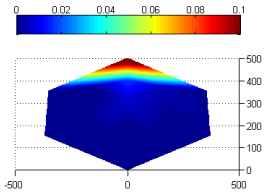
(a) image at 200° rotation
(b) GH precision at 200° rotation
(c) GH recall at 200° rotation
(d) GH efficiency at 200° rotation
(e) image at 325° rotation
(f) GH precision at 325° rotation
(g) GH recall at 325° rotation
(h) GH efficiency at 325° rotation
(i) image at 000° rotation
(j) GH precision at 000° rotation
(k) GH recall at 000° rotation
(l) GH efficiency at 000° rotation
(m) image at 035° rotation
(n) GH precision at 035° rotation
(o) GH recall at 035° rotation
(p) GH efficiency at 035° rotation
(q) image at 160° rotation
(r) GH precision at 160° rotation
(s) GH recall at 160° rotation
(t) GH efficiency at 160° rotation

Figure 119. Heat maps for descriptor GH in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

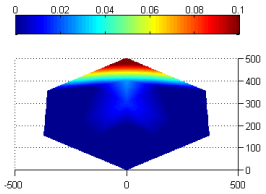
(a) image at 200° rotation



(b) GM precision at 200° rotation

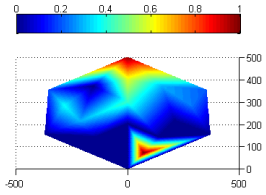


(c) GM recall at 200° rotation

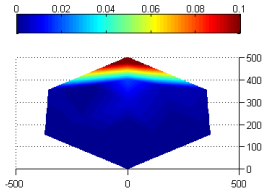


(d) GM efficiency at 200° rotation

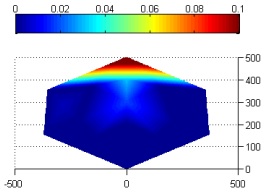
(e) image at 325° rotation



(f) GM precision at 325° rotation

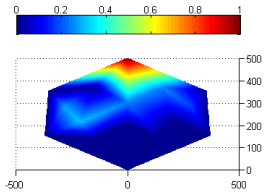


(g) GM recall at 325° rotation

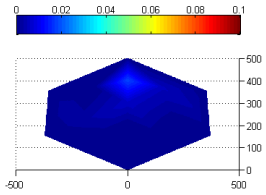


(h) GM efficiency at 325° rotation

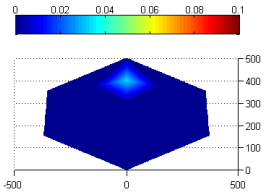
(i) image at 000° rotation



(j) GM precision at 000° rotation

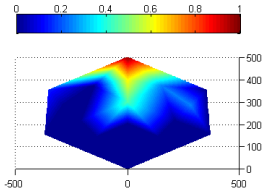


(k) GM recall at 000° rotation

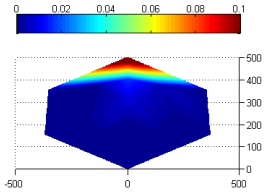


(l) GM efficiency at 000° rotation

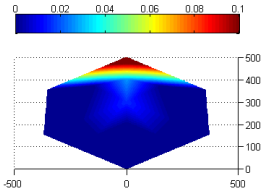
(m) image at 035° rotation



(n) GM precision at 035° rotation

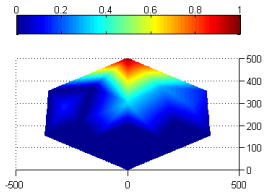


(o) GM recall at 035° rotation

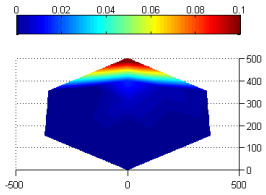


(p) GM efficiency at 035° rotation

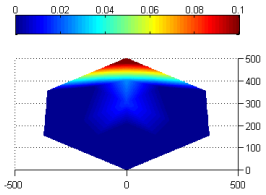
(q) image at 160° rotation



(r) GM precision at 160° rotation



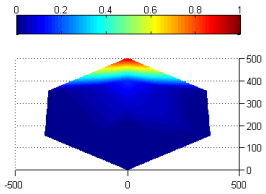
(s) GM recall at 160° rotation



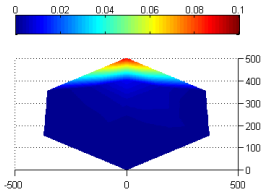
(t) GM efficiency at 160° rotation

Figure 120. Heat maps for descriptor GM in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

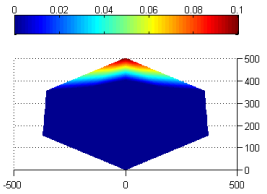
(a) image at 200° rotation



(b) DI precision at 200° rotation

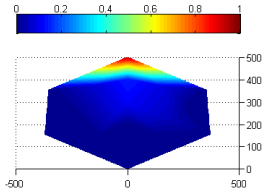


(c) DI recall at 200° rotation

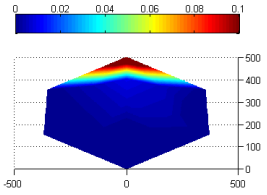


(d) DI efficiency at 200° rotation

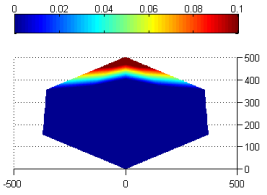
(e) image at 325° rotation



(f) DI precision at 325° rotation

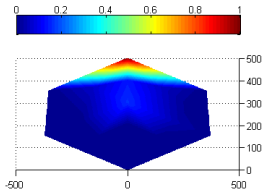


(g) DI recall at 325° rotation

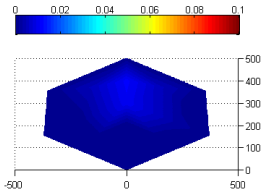


(h) DI efficiency at 325° rotation

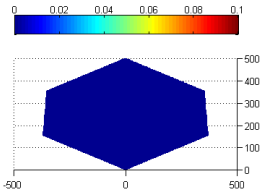
(i) image at 000° rotation



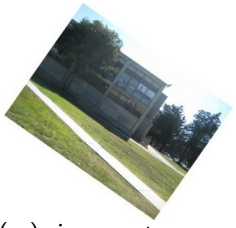
(j) DI precision at 000° rotation



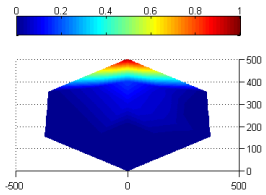
(k) DI recall at 000° rotation



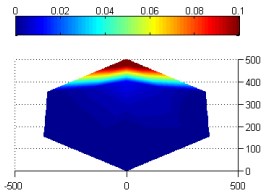
(l) DI efficiency at 000° rotation



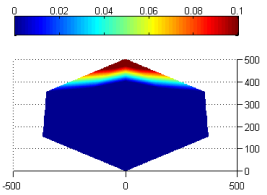
(m) image at 035° rotation



(n) DI precision at 035° rotation

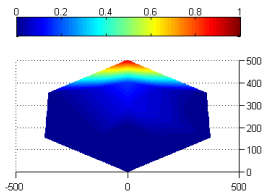


(o) DI recall at 035° rotation

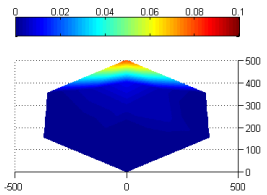


(p) DI efficiency at 035° rotation

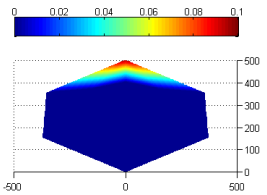
(q) image at 160° rotation



(r) DI precision at 160° rotation

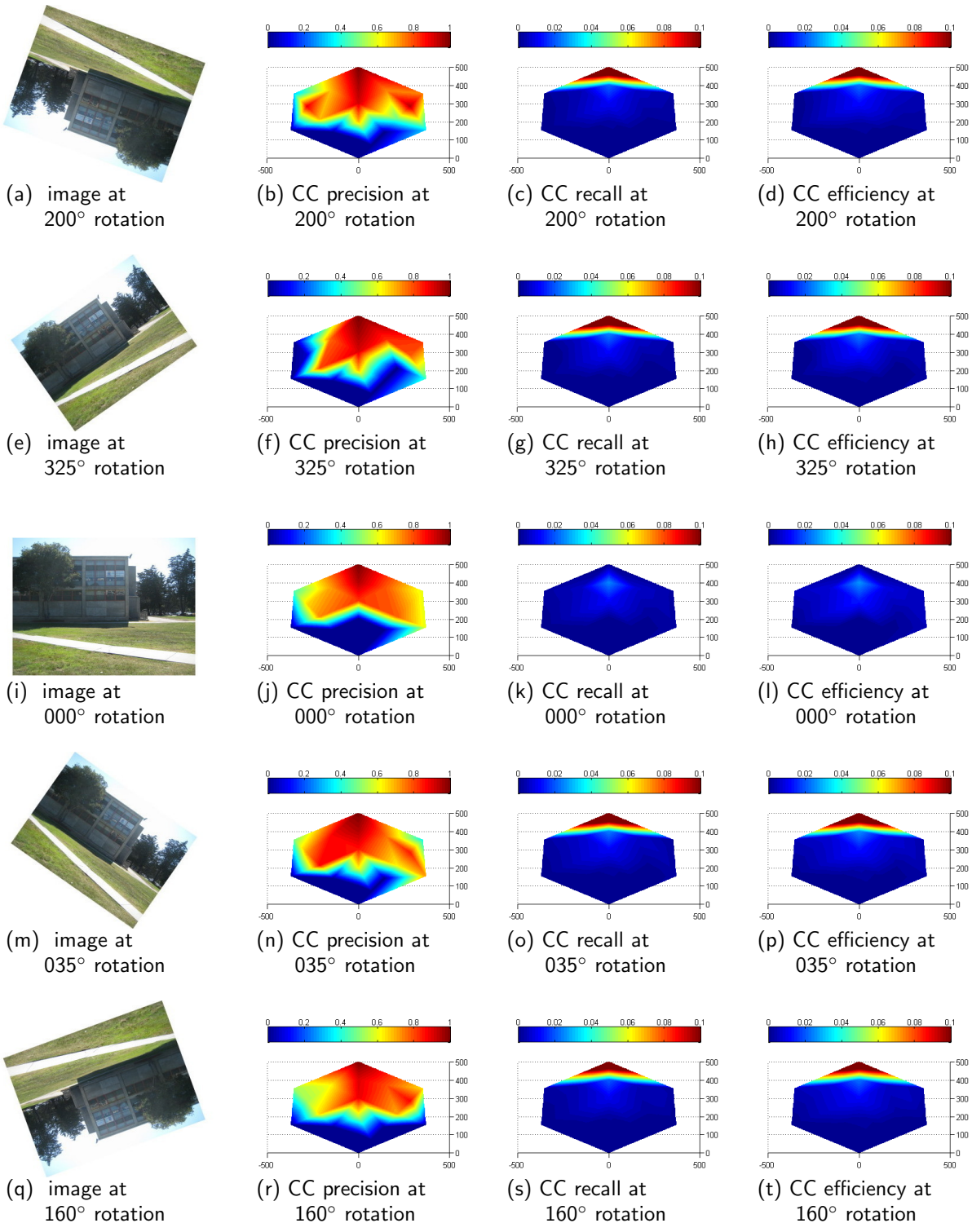


(s) DI recall at 160° rotation



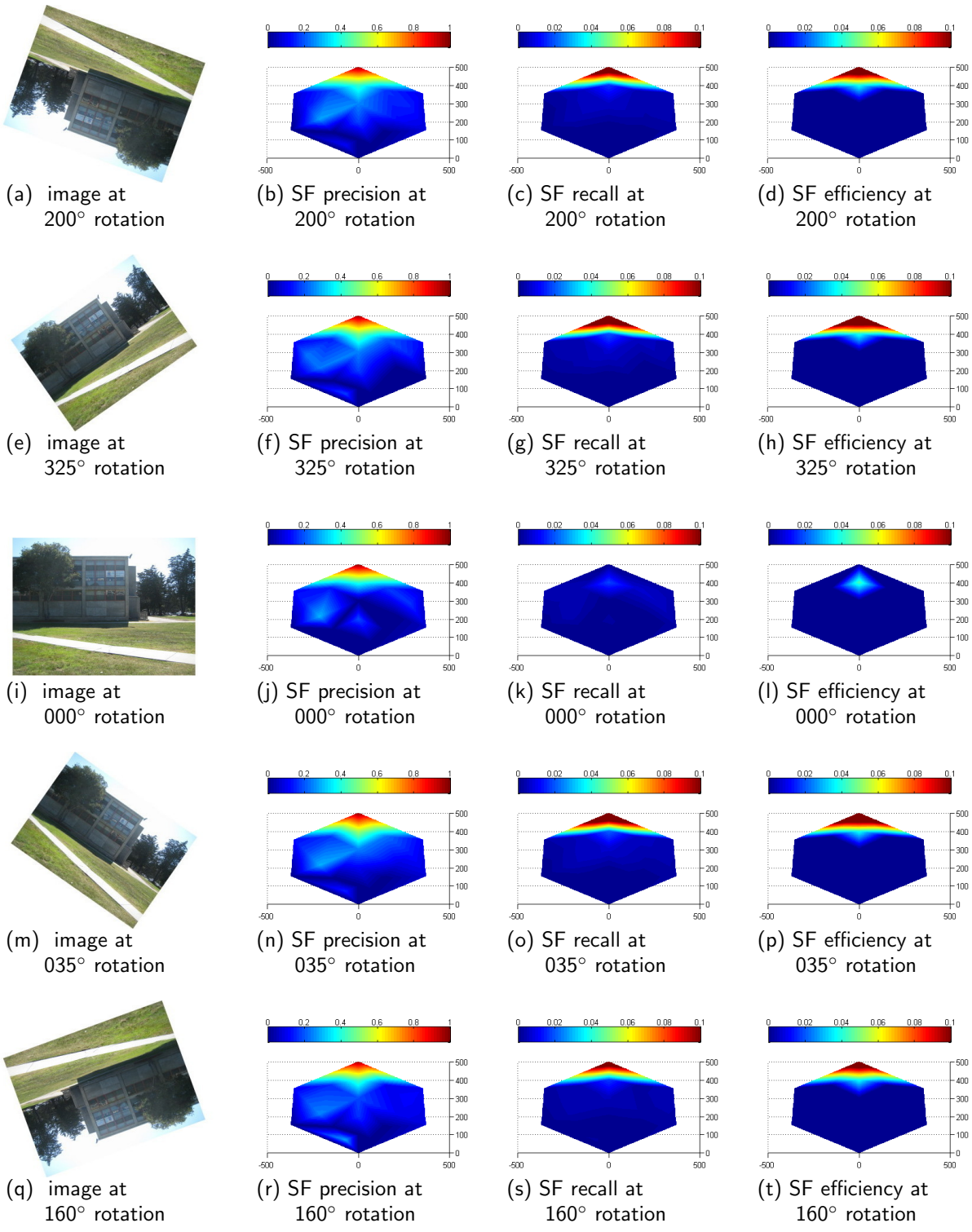
(t) DI efficiency at 160° rotation

Figure 121. Heat maps for descriptor DI in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



(a) image at 200° rotation
(b) CC precision at 200° rotation
(c) CC recall at 200° rotation
(d) CC efficiency at 200° rotation
(e) image at 325° rotation
(f) CC precision at 325° rotation
(g) CC recall at 325° rotation
(h) CC efficiency at 325° rotation
(i) image at 000° rotation
(j) CC precision at 000° rotation
(k) CC recall at 000° rotation
(l) CC efficiency at 000° rotation
(m) image at 035° rotation
(n) CC precision at 035° rotation
(o) CC recall at 035° rotation
(p) CC efficiency at 035° rotation
(q) image at 160° rotation
(r) CC precision at 160° rotation
(s) CC recall at 160° rotation
(t) CC efficiency at 160° rotation

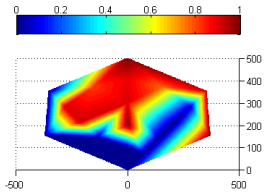
Figure 122. Heat maps for descriptor CC in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



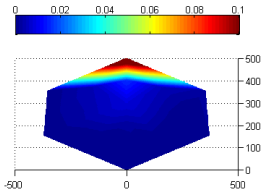
(a) image at 200° rotation
(b) SF precision at 200° rotation
(c) SF recall at 200° rotation
(d) SF efficiency at 200° rotation
(e) image at 325° rotation
(f) SF precision at 325° rotation
(g) SF recall at 325° rotation
(h) SF efficiency at 325° rotation
(i) image at 000° rotation
(j) SF precision at 000° rotation
(k) SF recall at 000° rotation
(l) SF efficiency at 000° rotation
(m) image at 035° rotation
(n) SF precision at 035° rotation
(o) SF recall at 035° rotation
(p) SF efficiency at 035° rotation
(q) image at 160° rotation
(r) SF precision at 160° rotation
(s) SF recall at 160° rotation
(t) SF efficiency at 160° rotation

Figure 123. Heat maps for descriptor SF in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

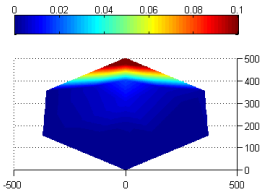
(a) image at 200° rotation



(b) PS precision at 200° rotation

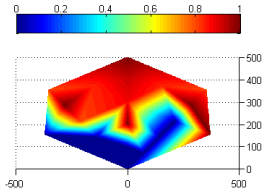


(c) PS recall at 200° rotation

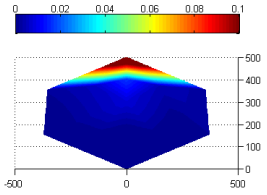


(d) PS efficiency at 200° rotation

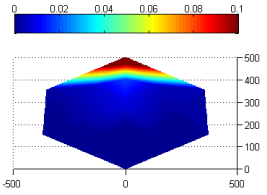
(e) image at 325° rotation



(f) PS precision at 325° rotation

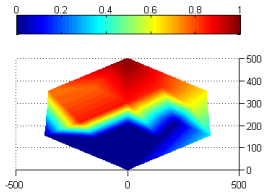


(g) PS recall at 325° rotation

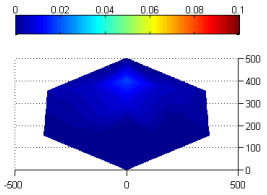


(h) PS efficiency at 325° rotation

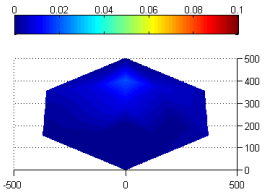
(i) image at 000° rotation



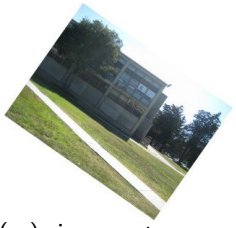
(j) PS precision at 000° rotation



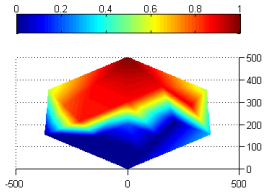
(k) PS recall at 000° rotation



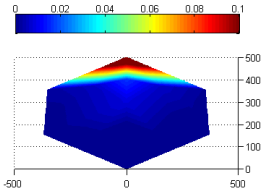
(l) PS efficiency at 000° rotation



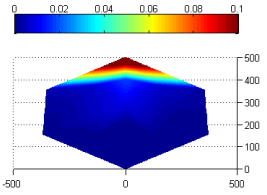
(m) image at 035° rotation



(n) PS precision at 035° rotation

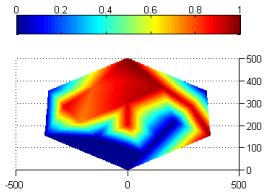


(o) PS recall at 035° rotation

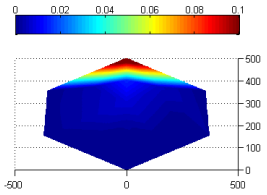


(p) PS efficiency at 035° rotation

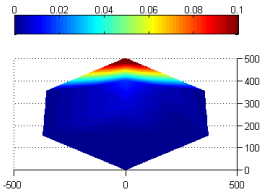
(q) image at 160° rotation



(r) PS precision at 160° rotation

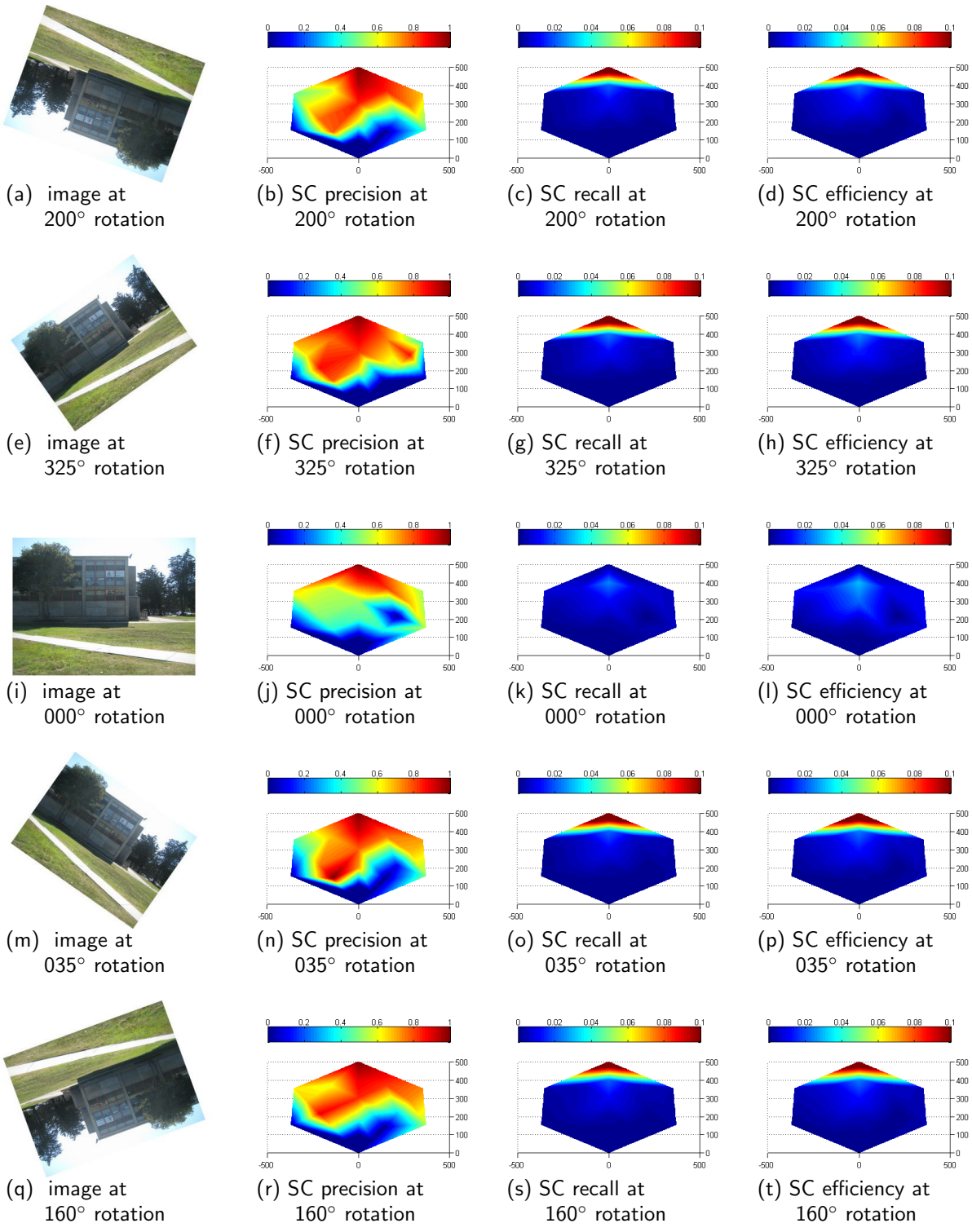


(s) PS recall at 160° rotation



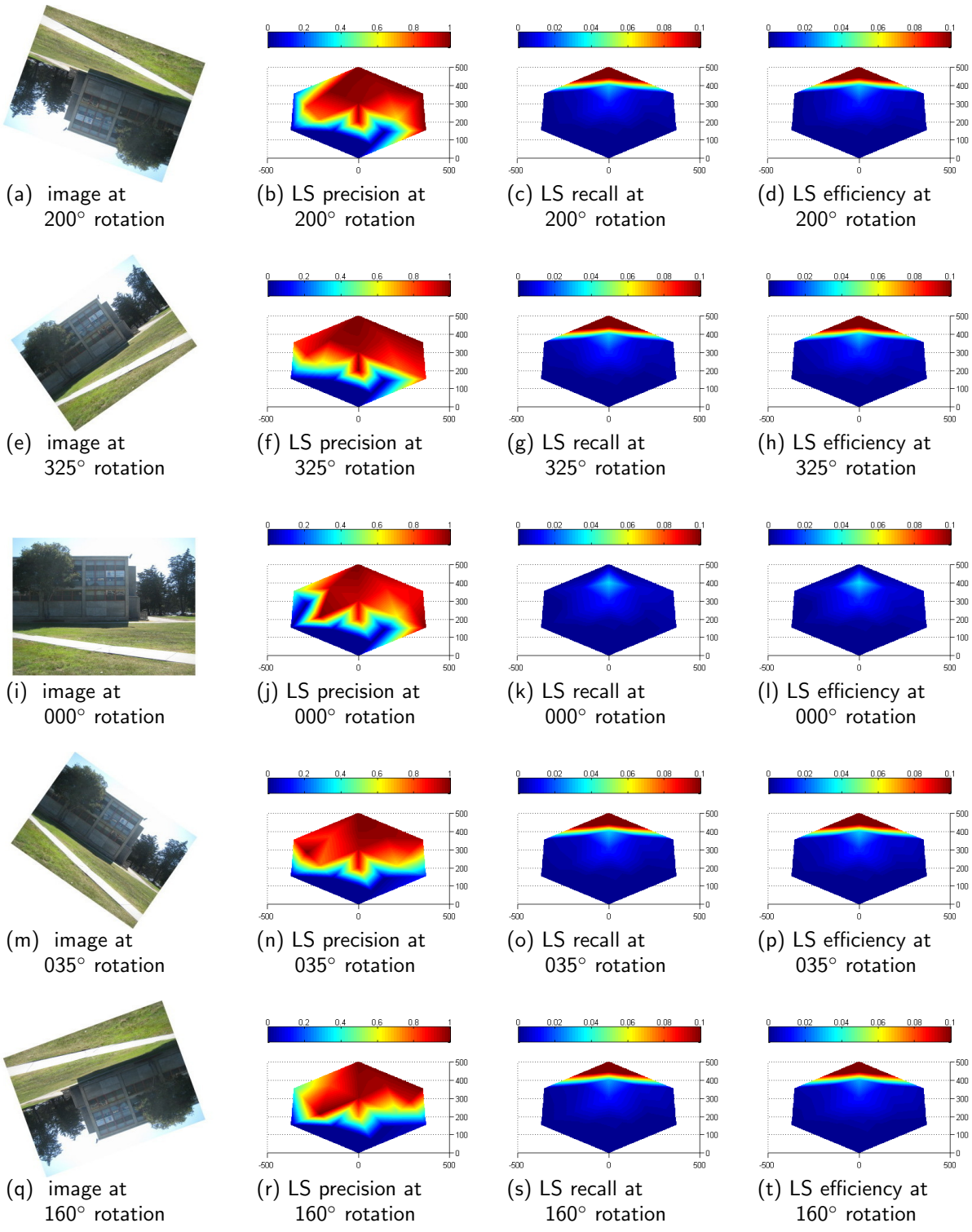
(t) PS efficiency at 160° rotation

Figure 124. Heat maps for descriptor PS in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



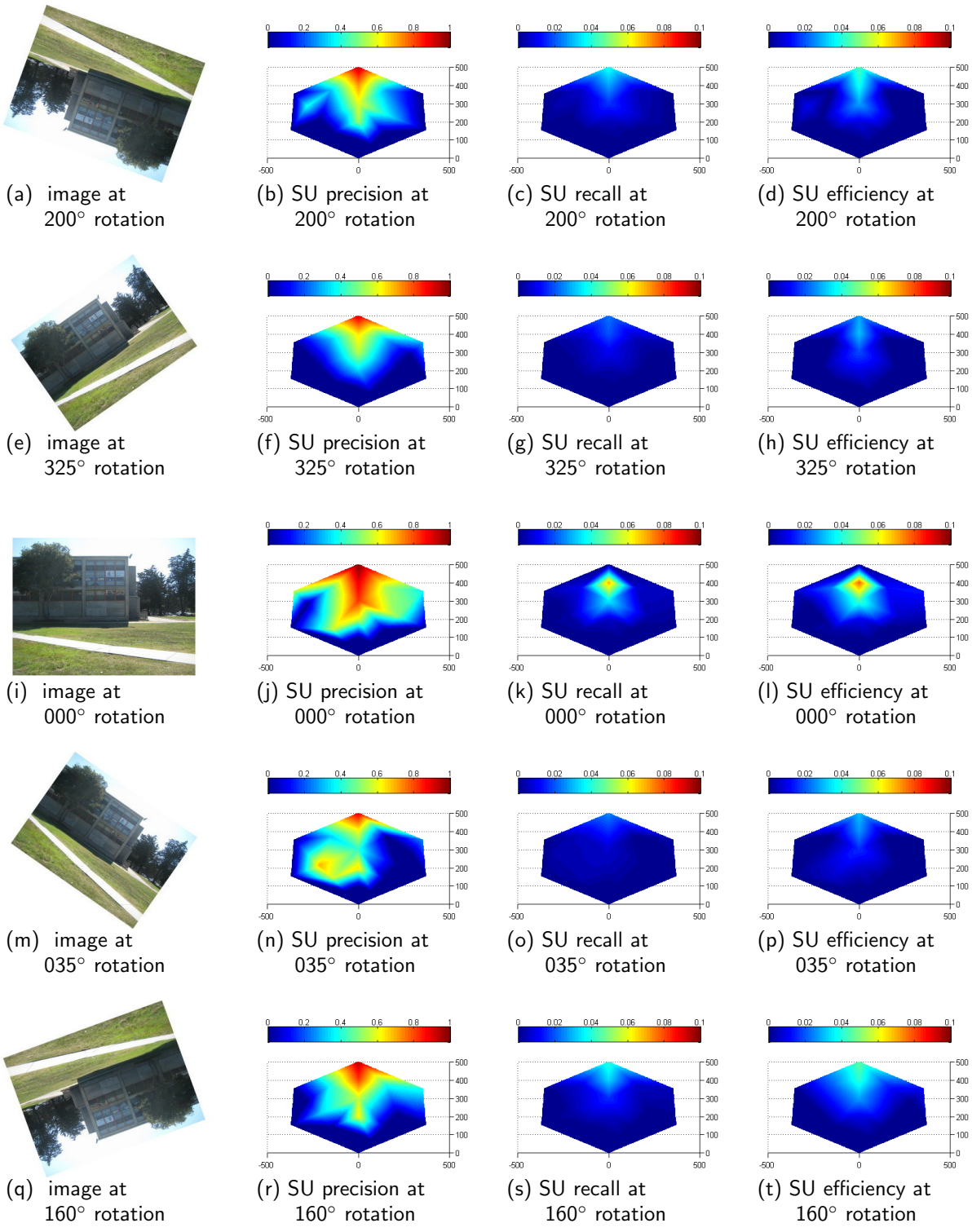
(a) image at 200° rotation
(b) SC precision at 200° rotation
(c) SC recall at 200° rotation
(d) SC efficiency at 200° rotation
(e) image at 325° rotation
(f) SC precision at 325° rotation
(g) SC recall at 325° rotation
(h) SC efficiency at 325° rotation
(i) image at 000° rotation
(j) SC precision at 000° rotation
(k) SC recall at 000° rotation
(l) SC efficiency at 000° rotation
(m) image at 035° rotation
(n) SC precision at 035° rotation
(o) SC recall at 035° rotation
(p) SC efficiency at 035° rotation
(q) image at 160° rotation
(r) SC precision at 160° rotation
(s) SC recall at 160° rotation
(t) SC efficiency at 160° rotation

Figure 125. Heat maps for descriptor SC in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



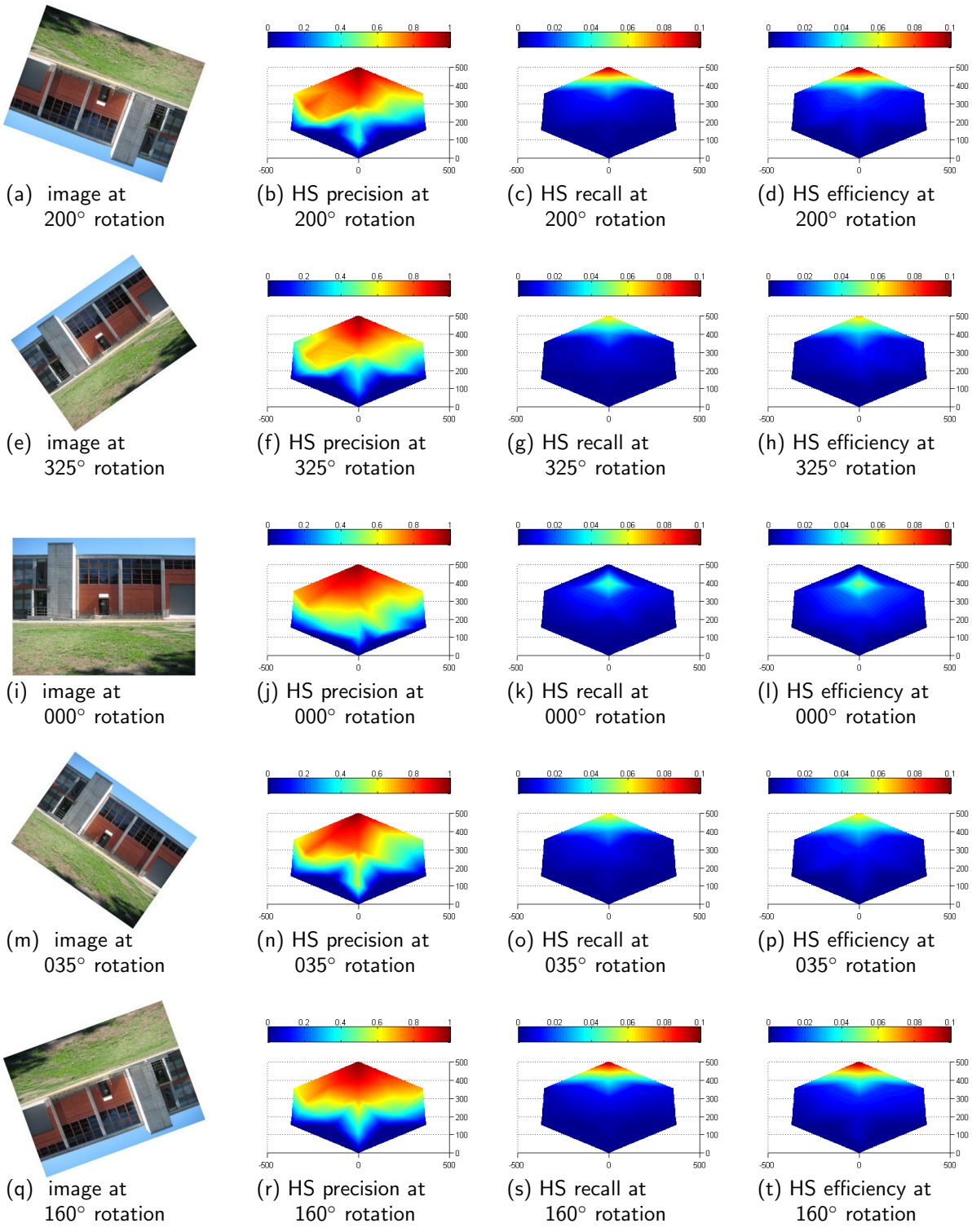
(a) image at 200° rotation
(b) LS precision at 200° rotation
(c) LS recall at 200° rotation
(d) LS efficiency at 200° rotation
(e) image at 325° rotation
(f) LS precision at 325° rotation
(g) LS recall at 325° rotation
(h) LS efficiency at 325° rotation
(i) image at 000° rotation
(j) LS precision at 000° rotation
(k) LS recall at 000° rotation
(l) LS efficiency at 000° rotation
(m) image at 035° rotation
(n) LS precision at 035° rotation
(o) LS recall at 035° rotation
(p) LS efficiency at 035° rotation
(q) image at 160° rotation
(r) LS precision at 160° rotation
(s) LS recall at 160° rotation
(t) LS efficiency at 160° rotation

Figure 126. Heat maps for descriptor LS in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



(a) image at 200° rotation
(b) SU precision at 200° rotation
(c) SU recall at 200° rotation
(d) SU efficiency at 200° rotation
(e) image at 325° rotation
(f) SU precision at 325° rotation
(g) SU recall at 325° rotation
(h) SU efficiency at 325° rotation
(i) image at 000° rotation
(j) SU precision at 000° rotation
(k) SU recall at 000° rotation
(l) SU efficiency at 000° rotation
(m) image at 035° rotation
(n) SU precision at 035° rotation
(o) SU recall at 035° rotation
(p) SU efficiency at 035° rotation
(q) image at 160° rotation
(r) SU precision at 160° rotation
(s) SU recall at 160° rotation
(t) SU efficiency at 160° rotation

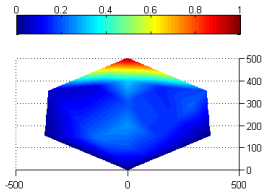
Figure 127. Heat maps for descriptor SU in the OutHaliganRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



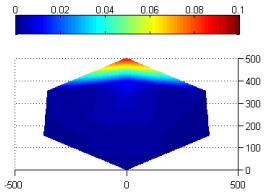
(a) image at 200° rotation
(b) HS precision at 200° rotation
(c) HS recall at 200° rotation
(d) HS efficiency at 200° rotation
(e) image at 325° rotation
(f) HS precision at 325° rotation
(g) HS recall at 325° rotation
(h) HS efficiency at 325° rotation
(i) image at 000° rotation
(j) HS precision at 000° rotation
(k) HS recall at 000° rotation
(l) HS efficiency at 000° rotation
(m) image at 035° rotation
(n) HS precision at 035° rotation
(o) HS recall at 035° rotation
(p) HS efficiency at 035° rotation
(q) image at 160° rotation
(r) HS precision at 160° rotation
(s) HS recall at 160° rotation
(t) HS efficiency at 160° rotation

Figure 128. Heat maps for descriptor HS in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

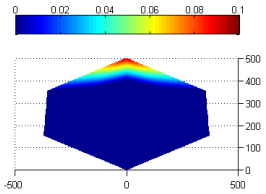
(a) image at 200° rotation



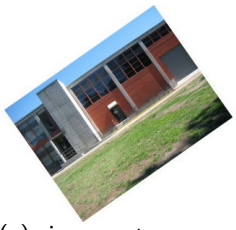
(b) CF precision at 200° rotation



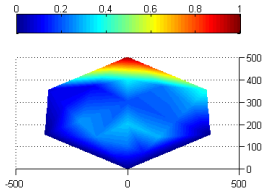
(c) CF recall at 200° rotation



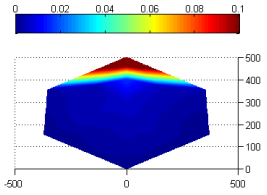
(d) CF efficiency at 200° rotation



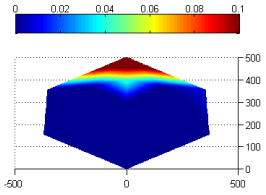
(e) image at 325° rotation



(f) CF precision at 325° rotation

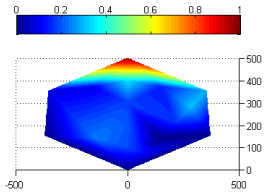


(g) CF recall at 325° rotation

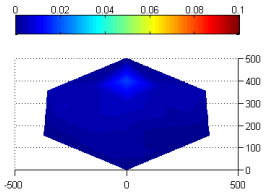


(h) CF efficiency at 325° rotation

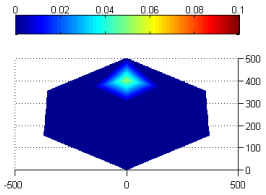
(i) image at 000° rotation



(j) CF precision at 000° rotation

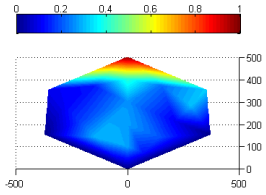


(k) CF recall at 000° rotation

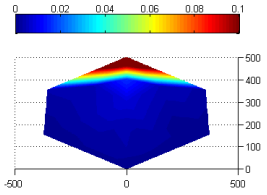


(l) CF efficiency at 000° rotation

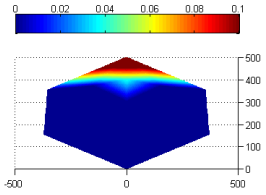
(m) image at 035° rotation



(n) CF precision at 035° rotation

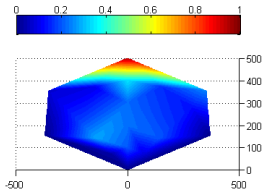


(o) CF recall at 035° rotation

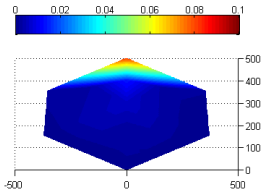


(p) CF efficiency at 035° rotation

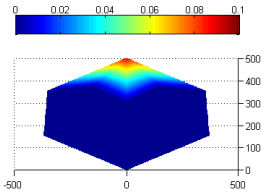
(q) image at 160° rotation



(r) CF precision at 160° rotation



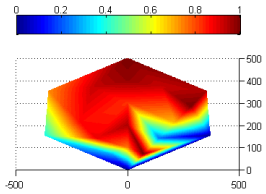
(s) CF recall at 160° rotation



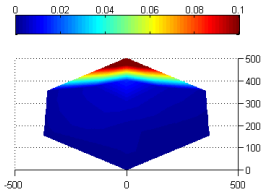
(t) CF efficiency at 160° rotation

Figure 129. Heat maps for descriptor CF in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

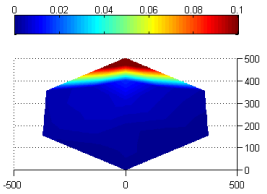
(a) image at 200° rotation



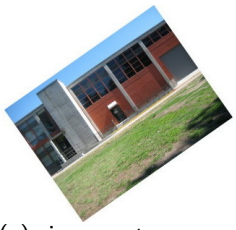
(b) GH precision at 200° rotation



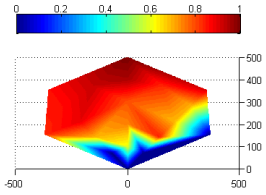
(c) GH recall at 200° rotation



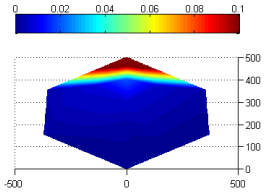
(d) GH efficiency at 200° rotation



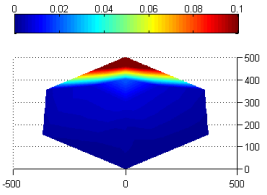
(e) image at 325° rotation



(f) GH precision at 325° rotation

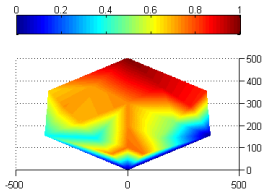


(g) GH recall at 325° rotation

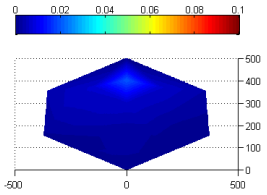


(h) GH efficiency at 325° rotation

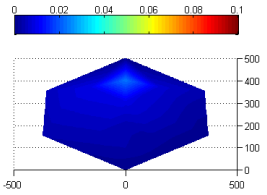
(i) image at 000° rotation



(j) GH precision at 000° rotation

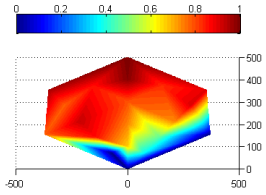


(k) GH recall at 000° rotation

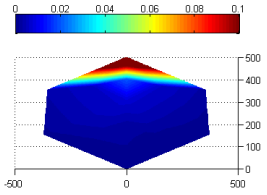


(l) GH efficiency at 000° rotation

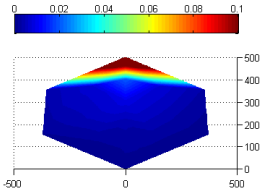
(m) image at 035° rotation



(n) GH precision at 035° rotation

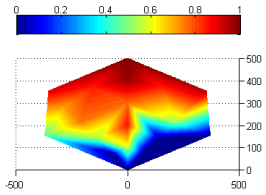


(o) GH recall at 035° rotation

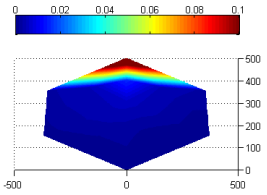


(p) GH efficiency at 035° rotation

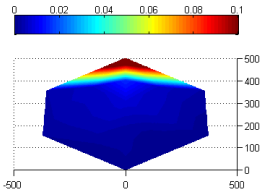
(q) image at 160° rotation



(r) GH precision at 160° rotation



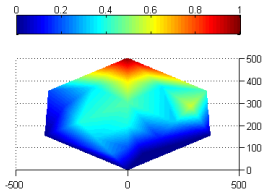
(s) GH recall at 160° rotation



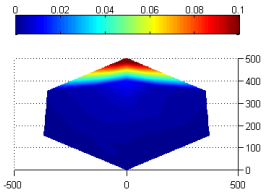
(t) GH efficiency at 160° rotation

Figure 130. Heat maps for descriptor GH in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

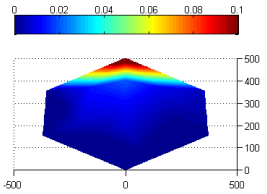
(a) image at 200° rotation



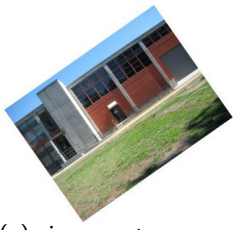
(b) GM precision at 200° rotation



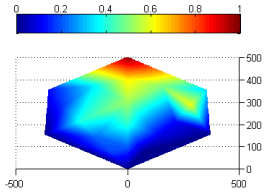
(c) GM recall at 200° rotation



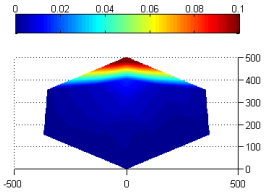
(d) GM efficiency at 200° rotation



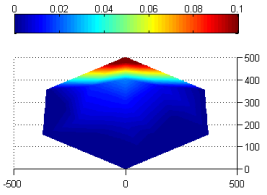
(e) image at 325° rotation



(f) GM precision at 325° rotation

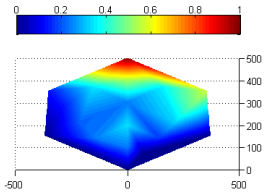


(g) GM recall at 325° rotation

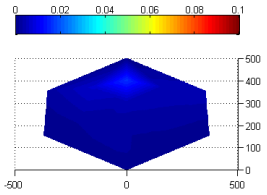


(h) GM efficiency at 325° rotation

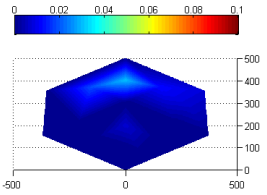
(i) image at 000° rotation



(j) GM precision at 000° rotation

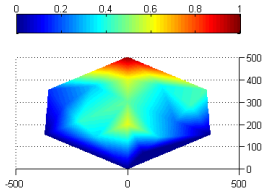


(k) GM recall at 000° rotation

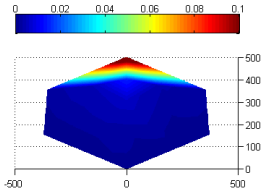


(l) GM efficiency at 000° rotation

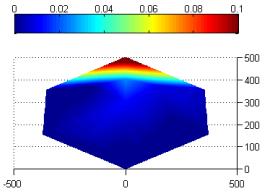
(m) image at 035° rotation



(n) GM precision at 035° rotation

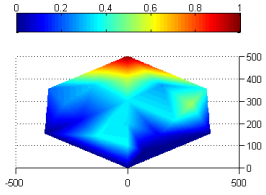


(o) GM recall at 035° rotation

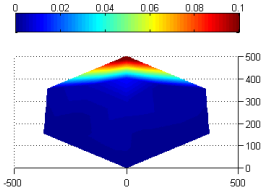


(p) GM efficiency at 035° rotation

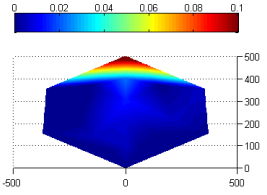
(q) image at 160° rotation



(r) GM precision at 160° rotation



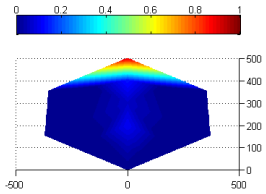
(s) GM recall at 160° rotation



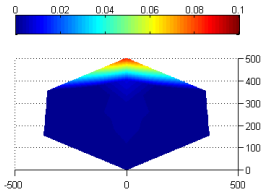
(t) GM efficiency at 160° rotation

Figure 131. Heat maps for descriptor GM in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

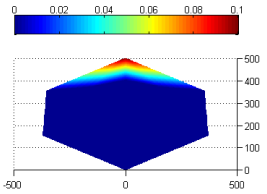
(a) image at 200° rotation



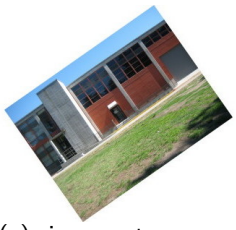
(b) DI precision at 200° rotation



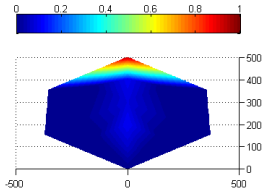
(c) DI recall at 200° rotation



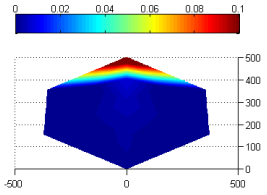
(d) DI efficiency at 200° rotation



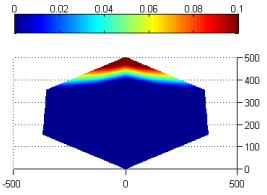
(e) image at 325° rotation



(f) DI precision at 325° rotation

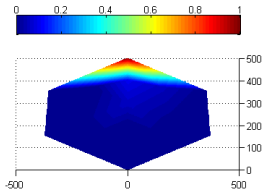


(g) DI recall at 325° rotation

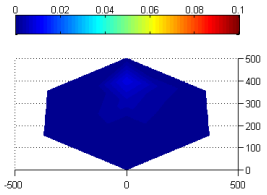


(h) DI efficiency at 325° rotation

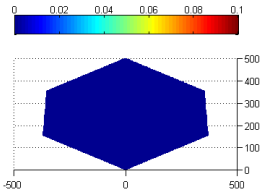
(i) image at 000° rotation



(j) DI precision at 000° rotation

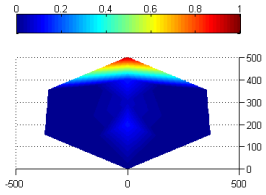


(k) DI recall at 000° rotation

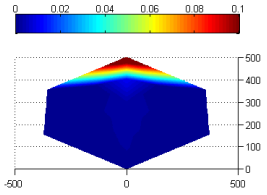


(l) DI efficiency at 000° rotation

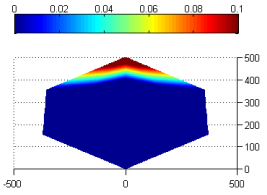
(m) image at 035° rotation



(n) DI precision at 035° rotation

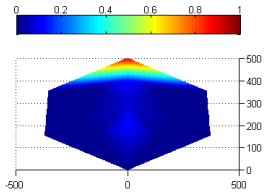


(o) DI recall at 035° rotation

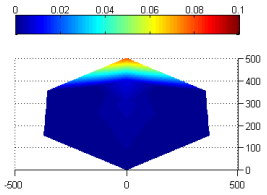


(p) DI efficiency at 035° rotation

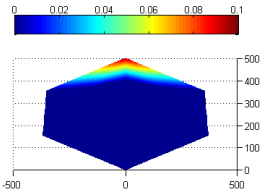
(q) image at 160° rotation



(r) DI precision at 160° rotation



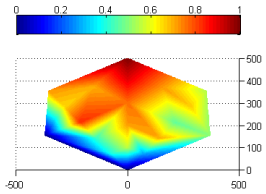
(s) DI recall at 160° rotation



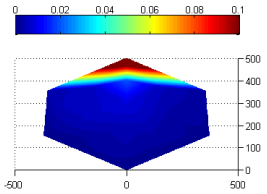
(t) DI efficiency at 160° rotation

Figure 132. Heat maps for descriptor DI in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

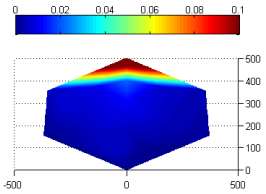
(a) image at 200° rotation



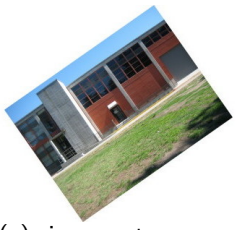
(b) CC precision at 200° rotation



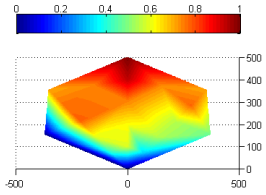
(c) CC recall at 200° rotation



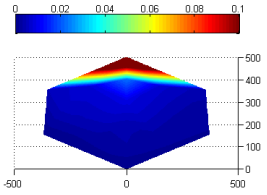
(d) CC efficiency at 200° rotation



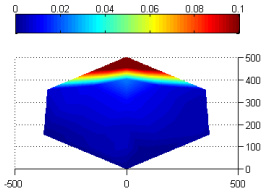
(e) image at 325° rotation



(f) CC precision at 325° rotation

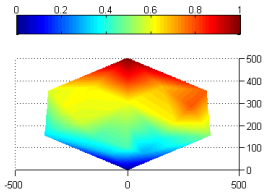


(g) CC recall at 325° rotation

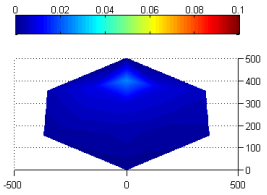


(h) CC efficiency at 325° rotation

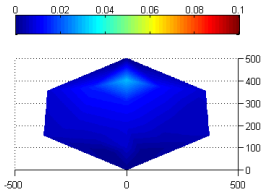
(i) image at 000° rotation



(j) CC precision at 000° rotation

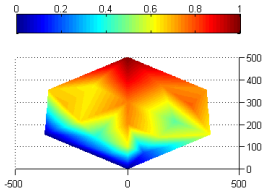


(k) CC recall at 000° rotation

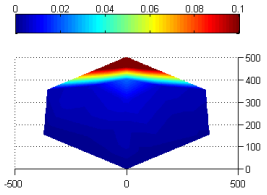


(l) CC efficiency at 000° rotation

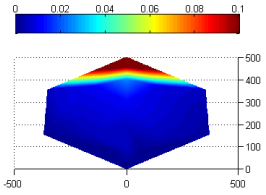
(m) image at 035° rotation



(n) CC precision at 035° rotation

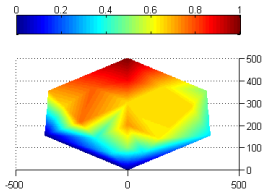


(o) CC recall at 035° rotation

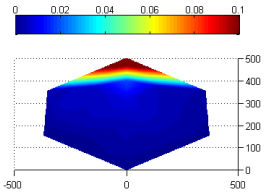


(p) CC efficiency at 035° rotation

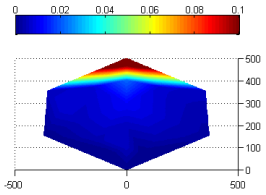
(q) image at 160° rotation



(r) CC precision at 160° rotation

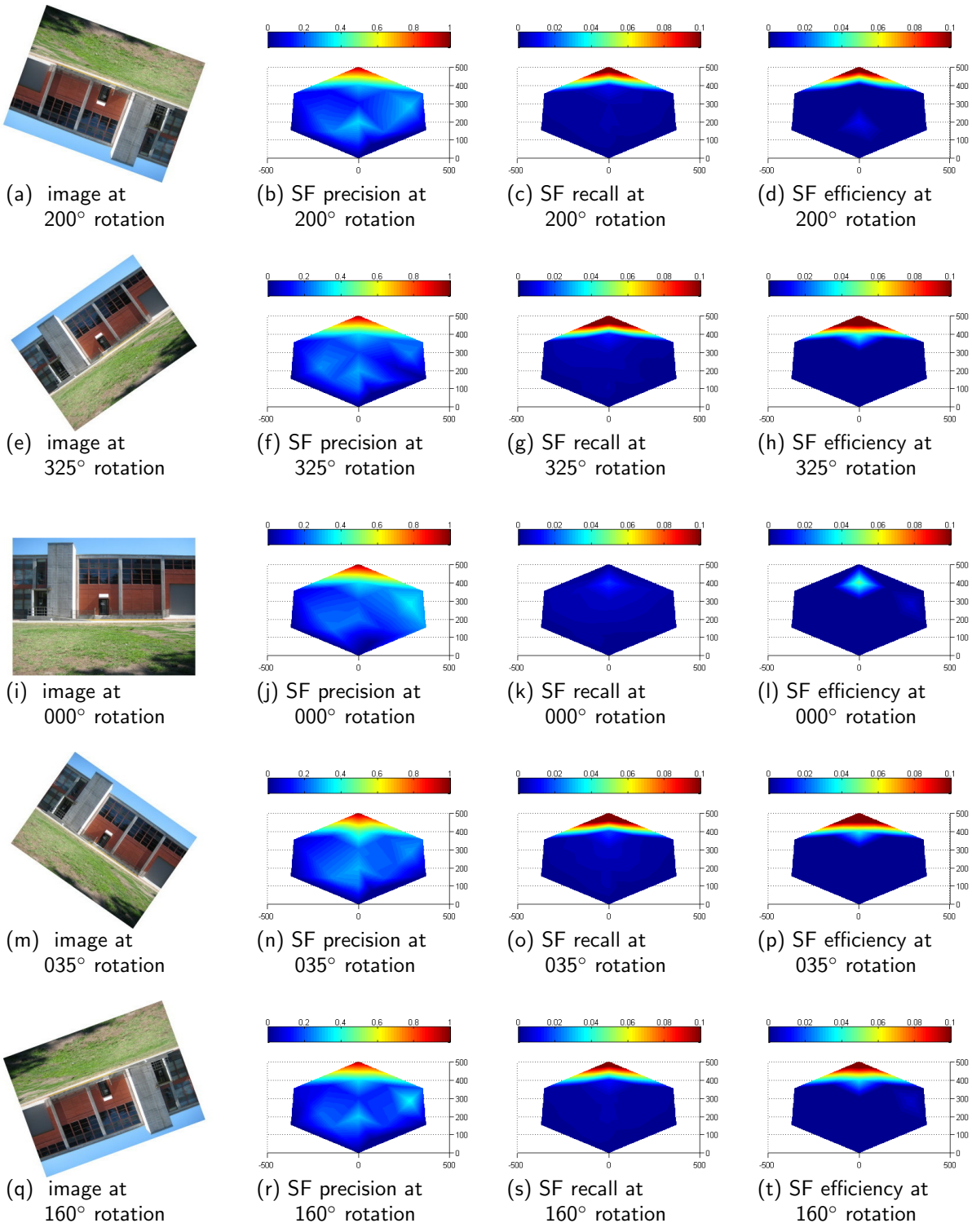


(s) CC recall at 160° rotation



(t) CC efficiency at 160° rotation

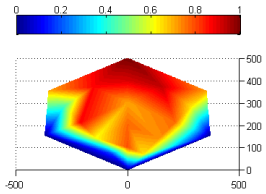
Figure 133. Heat maps for descriptor CC in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



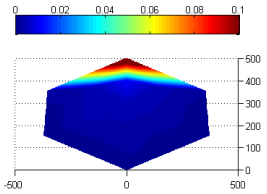
(a) image at 200° rotation
(b) SF precision at 200° rotation
(c) SF recall at 200° rotation
(d) SF efficiency at 200° rotation
(e) image at 325° rotation
(f) SF precision at 325° rotation
(g) SF recall at 325° rotation
(h) SF efficiency at 325° rotation
(i) image at 000° rotation
(j) SF precision at 000° rotation
(k) SF recall at 000° rotation
(l) SF efficiency at 000° rotation
(m) image at 035° rotation
(n) SF precision at 035° rotation
(o) SF recall at 035° rotation
(p) SF efficiency at 035° rotation
(q) image at 160° rotation
(r) SF precision at 160° rotation
(s) SF recall at 160° rotation
(t) SF efficiency at 160° rotation

Figure 134. Heat maps for descriptor SF in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

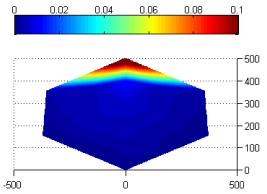
(a) image at 200° rotation



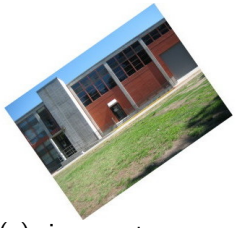
(b) PS precision at 200° rotation



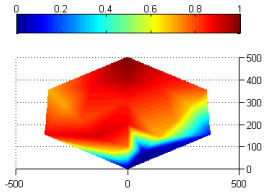
(c) PS recall at 200° rotation



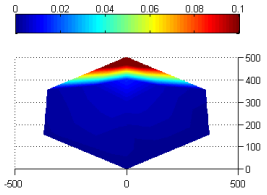
(d) PS efficiency at 200° rotation



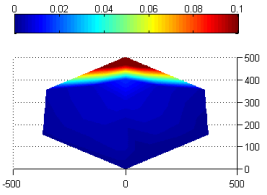
(e) image at 325° rotation



(f) PS precision at 325° rotation

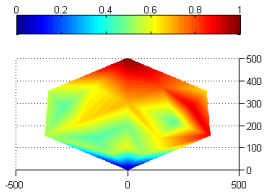


(g) PS recall at 325° rotation

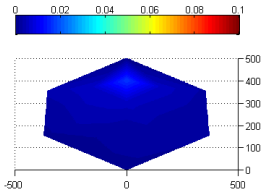


(h) PS efficiency at 325° rotation

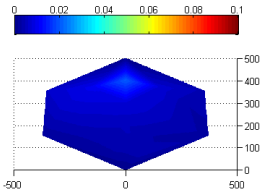
(i) image at 000° rotation



(j) PS precision at 000° rotation

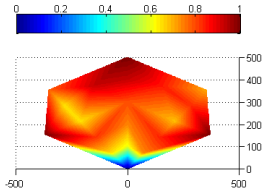


(k) PS recall at 000° rotation

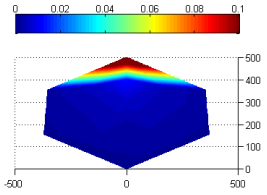


(l) PS efficiency at 000° rotation

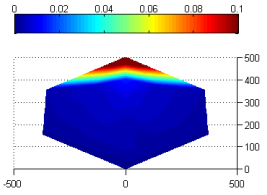
(m) image at 035° rotation



(n) PS precision at 035° rotation

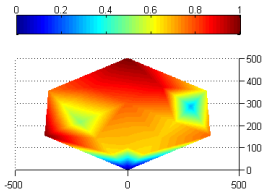


(o) PS recall at 035° rotation

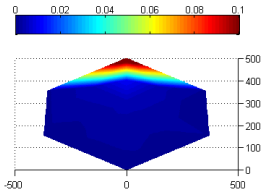


(p) PS efficiency at 035° rotation

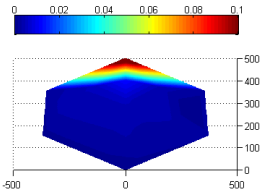
(q) image at 160° rotation



(r) PS precision at 160° rotation



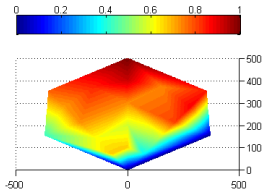
(s) PS recall at 160° rotation



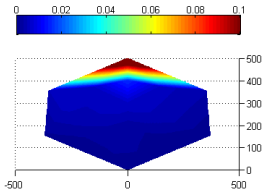
(t) PS efficiency at 160° rotation

Figure 135. Heat maps for descriptor PS in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

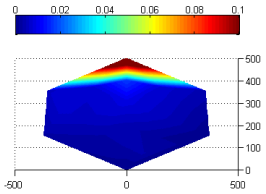
(a) image at 200° rotation



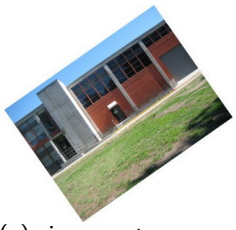
(b) SC precision at 200° rotation



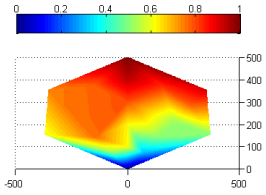
(c) SC recall at 200° rotation



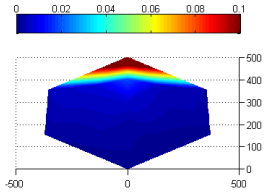
(d) SC efficiency at 200° rotation



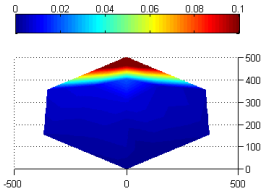
(e) image at 325° rotation



(f) SC precision at 325° rotation

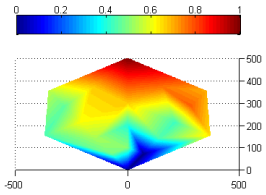


(g) SC recall at 325° rotation

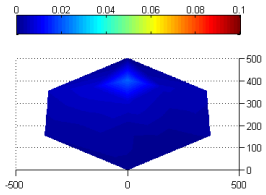


(h) SC efficiency at 325° rotation

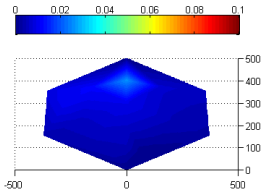
(i) image at 000° rotation



(j) SC precision at 000° rotation

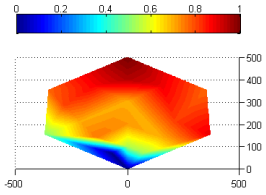


(k) SC recall at 000° rotation

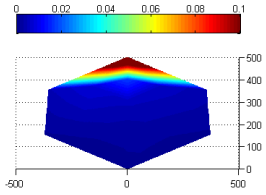


(l) SC efficiency at 000° rotation

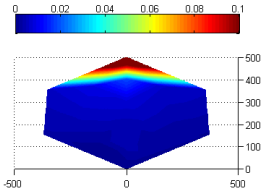
(m) image at 035° rotation



(n) SC precision at 035° rotation

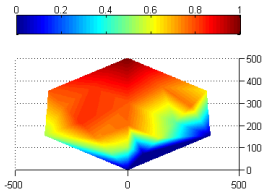


(o) SC recall at 035° rotation

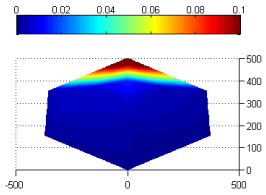


(p) SC efficiency at 035° rotation

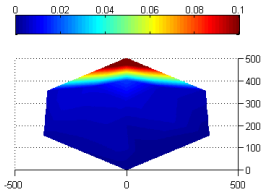
(q) image at 160° rotation



(r) SC precision at 160° rotation

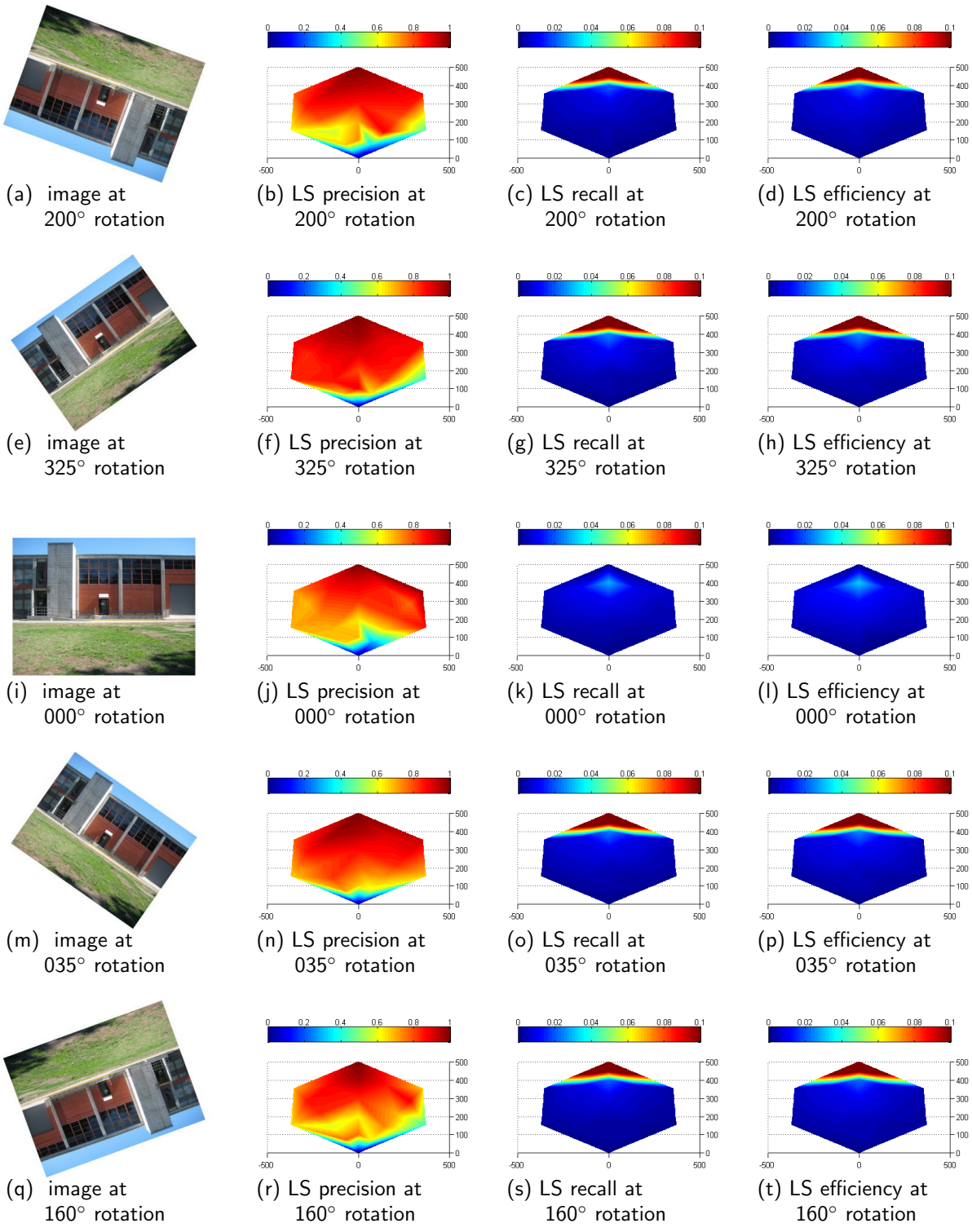


(s) SC recall at 160° rotation



(t) SC efficiency at 160° rotation

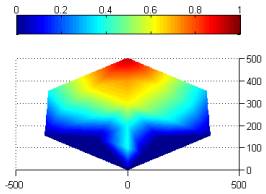
Figure 136. Heat maps for descriptor SC in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.



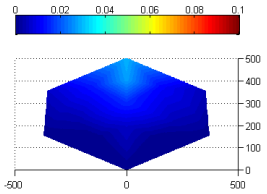
(a) image at 200° rotation
(b) LS precision at 200° rotation
(c) LS recall at 200° rotation
(d) LS efficiency at 200° rotation
(e) image at 325° rotation
(f) LS precision at 325° rotation
(g) LS recall at 325° rotation
(h) LS efficiency at 325° rotation
(i) image at 000° rotation
(j) LS precision at 000° rotation
(k) LS recall at 000° rotation
(l) LS efficiency at 000° rotation
(m) image at 035° rotation
(n) LS precision at 035° rotation
(o) LS recall at 035° rotation
(p) LS efficiency at 035° rotation
(q) image at 160° rotation
(r) LS precision at 160° rotation
(s) LS recall at 160° rotation
(t) LS efficiency at 160° rotation

Figure 137. Heat maps for descriptor LS in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

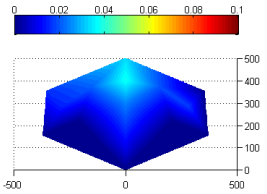
(a) image at 200° rotation



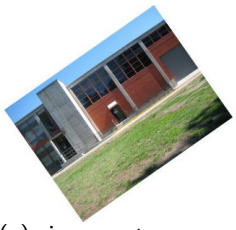
(b) SU precision at 200° rotation



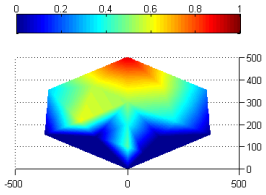
(c) SU recall at 200° rotation



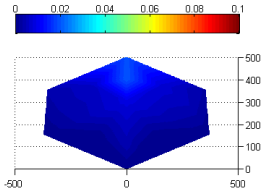
(d) SU efficiency at 200° rotation



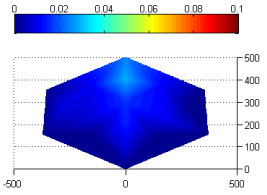
(e) image at 325° rotation



(f) SU precision at 325° rotation

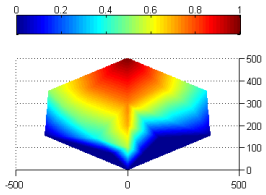


(g) SU recall at 325° rotation

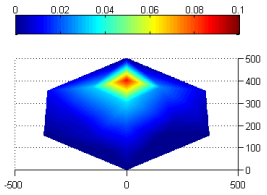


(h) SU efficiency at 325° rotation

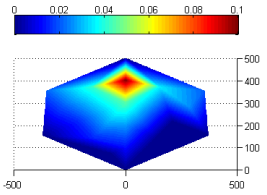
(i) image at 000° rotation



(j) SU precision at 000° rotation

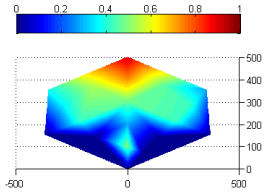


(k) SU recall at 000° rotation

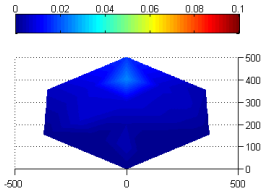


(l) SU efficiency at 000° rotation

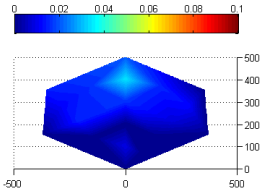
(m) image at 035° rotation



(n) SU precision at 035° rotation

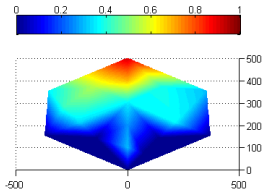


(o) SU recall at 035° rotation

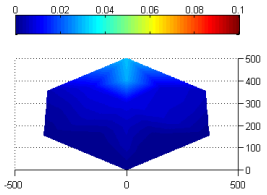


(p) SU efficiency at 035° rotation

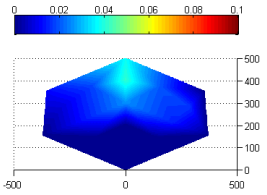
(q) image at 160° rotation



(r) SU precision at 160° rotation



(s) SU recall at 160° rotation



(t) SU efficiency at 160° rotation

Figure 138. Heat maps for descriptor SU in the OutUSLRef scene at various rotations. The hexagonal area represents the physical area in front of a target object in the scene located at (0,0). The reference image was taken from (0,500). The axis scale is 100=4m.

LIST OF REFERENCES

[1] S. Coren, L. M. Ward and J. T. Enns (1994). *Sensation and perception*. Harcourt Brace College Publishers, New York.

[2] Wurtz, R. and Kandel, E., Perception of motion, depth and form. In Kandell, E., Schwartz, J., and Messel, T., editors, Principles of Neural Science (4th edition), pp. 548-571.

[3] E. Sjoberg, “Autonomous Coordination And Online Motion Modeling For Mobile Robots," M.S. thesis, Naval Postgraduate School, Monterey, CA, 2007.

[4] P. Jensfelt, D. Kragic, J. Folkesson, M. Björkman “A framework for vision based bearing only 3D SLAM.” *Proceedings of IEEE International Conference on Robotics and Automation*, Orlando, FL, May 2006 (ICRA 2006).

[5] P. Saeedi, P. Lawrence, and D. Lowe, “Vision-Based 3D Trajectory Tracking for Unknown Environments” *International conference on Robotics and Automation,* Taipei, Taiwan, May 2003, pp. 1297-1303.

[6] S. Se, D. Lowe, and J. Little, “Vision-based Mobile Robot Localization and Mapping using Scale-Invariant Features” *Proceedings of IEEE International Conference on Robotics and Automation*, Seoul, Korea, May 2001 (ICRA 2001), pp. 2051-2058.

[7] P. Sala, R. Sim, A. Shokoufandeh, and S. Dickinson, “Landmark Selection for Vision-Based Navigation,” *IEEE Transactions on Robotics*, Vol. 22, No. 2, April 2006, pp. 334-349.

[8] N. Trawny, A.I. Mourikis, S.I. Roumeliotis, A. Johnson, J. Montgomery, A. Ansar, and L. Matthies, "Coupled Vision and Inertial Navigation for Pin-Point Landing" *In Proc. NASA Science Technology Conference* (NSTC'07), College Park, MD, June 19-21, 2007.

[9] B. Yamauchi, “Autonomous Urban Reconnaissance Using Man-Portable UGVs” *Proceedings of SPIE Vol. 6230: Unmanned Systems Technology VIII*, Orlando, FL, April 2006.

[10] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, “A Comparison of Affine Region Detectors.” *International Journal of Computer Vision*, Vol. 65, No. 1 of 2, 2005, pp. 43-72.

[11] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors" *IEEE Transactions on Pattern Analysis \& Machine Intelligence*, Vol. 27, No. 27, 2005, pp. 1615-1630.

[12] F. Fraundorfer and H. Bischof, “A novel performance evaluation method. of local detectors on non-planar scenes,” *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 3, No. 20-26, June 2005, pp. 33-33.

[13] J. Klippenstein and H. Zhang, “Quantitative Evaluation of Feature Extractors for Visual SLAM.” *In Proceedings of the Fourth Canadian Conference on Computer and Robot Vision*, May 2007, pp. 157-164.

[14] T. Tuytelaars and K. Mikolajczyk, "A Survey on Local Invariant Feature." *Course notes*, May 2006.

[15] Y. Itan, "Human Motion Perception Mechanism(s): A Functional Magnetic Resonance Imaging (fMRI) Experiment Applying Computational, Psychophysical and Physical Methods," 2005 Essay, CoMPLEX, University College London.

[16] C. Siagian, L. Itti, “Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention.” *PAMI*, Vol. 29, No. 2, February 2007, pp. 300-312.

[17] C. Carson, S. Belongie, S. Greenspan, J Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying." *IEEE Transactions on Pattern Analysis \& Machine Intelligence*, Vol. 24, No. 8, 2002, pp. 1026-1038.

[18] C. Harris and M. Stephens, "A combined corner and edge detector." *Avery Vision Conference,* 1988.

[19] W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features." *ISPRS Intercommission Workshop, Interlaken*, 1987.

[20] D. Lowe, "Distinctive image features from scale-invariant keypoints." *International Journal of Computer Vision*, Vol. 60, No. 2, November 2004, pp. 91-110.

[21] T. Lindeberg, "Feature detection with automatic scale selection." *International Journal of Computer Vision*, Vol. 30, No. 2, 1998, pp. 79-116.

[22] J. Matas, O. Chum, M. Urba, and T. Pajdla. "Robust wide baseline stereo from maximally stable extremal regions." Proc. of British Machine Vision Conference, pp. 384-396, 2002.

[23] A. J. Davison and D. W. Murray, ”Simultaneous localization and mapbuilding using active vision.” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24 No. 7, pp. 865-880, 2002.

[24] G. Tian, D. Gledhill, and D. Taylor, “Comprehensive interest points based imaging mosaic.” *Pattern Recognition Letters*, June 2003, pp. 1171-1179.

[25] J. Sivic and A. Zisserman, "Video Google: A text Retrieval Approach to Object Matching in Videos." *Proceedings of the Ninth International Conference on Computer Vision*, 2003, pp. 1470-1478.

[26] G. Dorko and C. Schmid, "Selection of Scale-Invariant Parts for Object Class Recognition." *Proceedings of the Ninth International Conference on Computer Vision*, 2003, pp. 634-640.

[27] R. Fergus, P. Perona and A Zisserman, "Object Class Recognition by Unsupervised Scale-Invariant Learning." *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2003, pp. 264-271.

[28] S. Lazebnik, C. Schmid and J. Ponce, "Sparse Texture Representation Using Affine-Invariant Neighborhoods." *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2003, pp. 319-324.

[29] T. Tuytelaars and L. Van Gool, "Matching Widely Separated Views Based on Affine Invariant Regions." *International Journal on Computer Vision*, Vol. 1, No. 59, 2004, pp. 61-85.

[30] R. Harrell, D. Slaughter and P. Adsit, "A fruit-tracking system for robotic harvesting." *MVA*, Vol. 2, No. 2, 1989, pp. 69-80.

[31] D. Lowe, “Object recognition from local scale-invariant features.” *In International Conference on Computer Vision,* Corfu, Greece, 1999, pp. 1150-1157.

[32] S. Thrun et al., “MINERVA: A second-generation museum tour-guide robot,” *in Proc. IEEE Int. Conf. on Robotics and automation, Detroit, MI*, Vol. 3, 1999, pp. 1999-2005.

[33] C. Schmid, R. Mohrand and C. Bauckhage, “Evaluation of Interest Point Detectors,” *Int'l Journal of Computer Vision*, Vol. 37, No. 2, 151-172, 2000.

[34] C. Harris and M.J. Stephens. “A combined corner and edge detector.” *In Alvey Vision Conference*, 1988, pp. 147-152.

[35] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, 2004, pp. 91-110.

[36] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework." *International Journal of Computer Vision*, Vol. 56, No. 3, March 2004, pp. 221-255.

[37] H. Bingrong, L. Maohai, L. Ronghua, “Novel Mobile Robot Simultaneous Loclization and Mapping Using Rao-Blackwellised Particle Filter.” *International Journal of Advanced Robotic Systems*, September 2006, Vol. 3, Issue 3, pp. 231-238.

[38] I.T. Jolliffe, “Principal Component Analysis.” *Springer series in statistics*, Springer-Verlag, 1986.

[39] Y. Ke and R. Sukthankar. “PCA-SIFT: A more distinctive representation for local image descriptors.” *In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR),* 2004.

[40] H. Bay, T. Tuytelaars and L. Van Gool, “SURF: Speeded Up Robust Features.” *Proceedings of the 9th European Conference on Computer Vision*, 2006, Springer LNCS, Vol. 3951, Issue 1, pp. 404-417.

[41] S. Belongie, J. Malik, and J. Puzicha. “Shape matching and object recognition using shape contexts.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp. 509-522, 2002.

[42] J. Canny. “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 8, No. 6, pp. 679–698, 1986.

[43] L. Van Gool, T. Moons, and D. Ungureanu. “Affine / photometric invariants for planar intensity patterns.” In *Proceedings of the 4th European Conference on Computer Vision*, Cambridge, UK, pp. 642-651, 1996.

[44] W. Freeman and E. Adelson. “The design and use of steerable filters.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 9, pp. 891-906, 1991.

[45] J. Koenderink and A. van Doorn. “Representation of local geometry in the visual system.” Biological Cybernetics, Vol. 55, pp. 367-375, 1987.

[46] F. Schaffalitzky and A. Zisserman. “Multi-view matching for unordered image sets.” In *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, pp. 414-431, 2002.

[47] T. Jebara, A. Azarbayejani, and A. Pentland, “3D structure from 2D motion: The inverse Hollywood problem: Getting models and motion out of video for post-production, MPEG, and more.” In *IEEE Signal Processing Magazine*, pp. 66-84, 1999.

[48] M. Tomono, “3D Localization and Mapping Using a Single Camera Based on Structure-from-Motion with Automatic Baseline Selection.” *IEEE Conf. on Robotics and Automation*. 2005, pp. 3342-3347.

[49] J. Shi and C. Tomasi, “Good features to track.” *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593-600.

[50] C. Tomasi, T. Kanade. “Detection and tracking of point features.” *Tech. Rept. CMU-CS-91132*. Pittsburgh: Carnegie Mellon U. School of Computer Science, 1991.

[51] J. Meltzer, M. H. Yang, R. Gupta, and S. Soatto. "Multiple view feature descriptors from image sequences via kernel principal component analysis." *Computer Vision - ECCV 2004*, Pt 1. 3021, pp. 215-227.

[52] T. Jebara, A. Azarbayejani, and A. Pentland. “3D structure from 2D motion: The inverse Hollywood problem: Getting models and motion out of video for post-production, MPEG, and more.” *IEEE Signal Processing Magazine*, May 1999, pp. 66-84.

[53] Durrant-whyte, H.F., Leonard, J.J. "Simultaneous map building and localization for an autonomous mobilerobot." *Intelligent Robots and Systems' 91.'Intelligence for Mechanical Systems, Proceedings IROS'91*. 1991, pp. 1442-1447.

[54] E. Eade, & T. Drummond, “Scalable monocular SLAM.” *Computer Vision and Pattern Recognition (CVPR)*, New York, NY, 2006, pp. 469-476.

[55] Montemerlo, S Thrun, D Koller, B Wegbreit, "FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem," in *Proceedings of the American Association for Artificial Intelligence (AAAI) National Conference on Artificial Intelligence*. Edmonton, Canada, 2002.

[56] A. J. Davison, and N. D. Molton, “MonoSLAM: Real-Time Single Camera SLAM.” *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. 29, Issue 6, June 2007, pp. 1052-1067.

[57] M. Israël, E. L. van den Broek, P. van der Putten, and M. J. den Uyl, “Real time automatic scene classification,” in *Proceedings of the Sixteenth Belgium-Netherlands Artificial Intelligence Conference*, R. Verbrugge, N. Taatgen, and L. R. B. Schomaker, Eds., 2004, pp. 401–402.

[58] N. Rasiwasia and N. Vasconcelos, "Scene Classification with Low-dimensional Semantic Spaces and Weak Supervision." *In IEEE Conference on Computer Vision and Pattern Recognition,* Anchorage, pp. 1-6, 2008.

[59] A. Bosch, A. Zisserman, and X. Munoz. “Scene classification via pLSA.” In *ECCV*, pp. 517–30, Graz, Austria, 2006.

[60] G. Csurka, C. Bray, C. Dance, and L. Fan. “Visual categorization with bags of keypoints.” Workshop on Statistical Learning in Computer Vision,” In *ECCV*, pp. 1–22, 2004.

[61] F.-F. Li and P. Perona. “A bayesian hierarchical model for learning natural scene categories.” In *IEEE CVPR*, pp. 524–531, 2005.

[62] J. Liu and M. Shah. “Scene modeling using co-clustering.” in *ICCV*, 2007.

[63] E. Nowak, F. Jurie, and B. Triggs. “Sampling strategies for bag-of-features image classification.” In *Proc. ECCV*, Vol. 4, pp. 490–503, 2006.

[64] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool. “Modeling scenes with local descriptors and latent aspects.” In *ICCV*, Vol. 1, pp. 883–90, 2005.

[65] N. W. Campbell, W. P. J. Mackeown, B. T. Thomas, and T. Troscianko. “Interpreting image databases by region classification,” *Pattern Recognition*, Vol. 30 No. 4, pp. 555–563, 1997.

[66] B. E. Boser, I. M. Guyon, and V. N. Vapnik. “A training algorithm for optimal margin classifiers.” In *D. Haussler, editor, 5th Annual ACM Workshop on COLT*, pp. 144-152, Pittsburgh, PA, 1992.

[67] R. Hess, MSVC++ SIFT source code. Oregon State University, OR, [Online] Available: http://web.engr.oregonstate.edu/~hess/index.html. [Accessed: Nov 2, 2007].

[68] OpenCV. Intel, [Online] Available: http://www.intel.com/research/mrl/research /opencv/. [Accessed: Jun 17, 2007].

[69] GSL, [Online], Available: http://www.gnu.org/software/software.html, [Accessed: Nov 2, 2007].

[70] Affine Covariant Region Evaluation Software. Mikolajczyk, [Online] Available: http://www.robots.ox.ac.uk/~vgg/research/affine/.

[71] M. A. Fischler, R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Communications of the ACM*, Vol. 24, pp. 381-395, 1981.

[72] R.I. Hartley and A. Zisserman. “*Multiple View Geometry in Computer Vision*,” second ed. Cambridge Univ. Press, 2004.

[73] D. Forsyth, and J. Ponce, "Computer Vision: A Modern Approach," Upper Saddle River, N.J. ; London: Prentice Hall, 2003.

[74] S. Thrun, W. Burgard, and D. Fox, “FastSLAM algorithm. Probabilistic robotics.” *Massachusetts Institute of Technology Press*, Cambridge, Massachusetts, 2005, pp. 437-483.

[75] M. McVicker, “Effects Of Different Camera Motions On The Error In Estimates Of Epipolar Geometry Between Two Dimensional Images In Order To Provide A Framework For Solutions To Vision Based Simultaneous Localization And Mapping (SLAM)," M.S. thesis, Naval Postgraduate School, Monterey, CA, 2007.

[76] J. P. Lewis, "Fast normalized cross-correlation," in *Vision Interface. Canadian Image Processing and Pattern Recognition Society*, 1995, pp. 120-123.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California